**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61K

A2

(11) International Publication Number: **WO 98/27932**

(43) International Publication Date: 2 July 1998 (02.07.98)

(21) International Application Number: PCT/US97/23518

(22) International Filing Date: 18 December 1997 (18.12.97)

(30) Priority Data:
60/033,870 20 December 1996 (20.12.96) US

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US).

(74) Agents: STEFFE, Eric, K.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005-3934 (US) et al.

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*Without international search report and to be republished upon receipt of that report.*

(54) Title: CEREBELLUM AND EMBRYO SPECIFIC PROTEIN

(57) Abstract

The present invention relates to a novel cerebellum and embryo specific (CESP) protein which is a member of the myocardial factor superfamily. In particular, isolated nucleic acid molecules are provided encoding the human CESP protein. CESP polypeptides are also provided as are vectors, host cells and recombinant methods for producing the same.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# Cerebellum and Embryo Specific Protein

## *Background of the Invention*

### *Field of the Invention*

The present invention relates to a novel endothelial factor. More specifically, isolated nucleic acid molecules are provided encoding a human cerebellum and embryo specific protein. Cerebellum and embryo specific polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are diagnostic and therapeutic methods relating to cerebellum and embryo specific protein-related disorders.

### *Related Art*

Myocardial necrosis results from occlusion of a coronary artery by a thrombus, which forms on a destabilized atherosclerotic plaque, often following plaque rupture. Plaques most prone to rupture and thrombosis may initially be only mildly stenotic (i.e., 50% to 60% stenotic). However, myocardial damage proceeds rapidly as a "wave front" of injury, moving from endocardium to epicardium and may become complete and irreversible within three to four hours, unless the infarct zone is adequately nourished by collateral blood supply or unless recanalization of the artery (i.e., revascularization) is accomplished. See Rogers, W.J., *Am. J. Med. 99*:195-206 (1995). However, collateral circulation typically doesn't develop until a severe coronary artery stenosis has already developed (Schaper, W., *European Heart J. 16*:66-68 (1995)).

In one model of coronary angiogenesis, vascular formation occurs through three major stages including 1) vessel dilation and endothelial cell activation; 2) formation of a new vascular channel; and 3) maturation of the new vessel and final differentiation of all vascular cells (Rakusan, K., Coronary Angiogenesis: From Morphometry to Molecular Biology and Back, in: Claycomb, W.C and Di Nardo, P., eds., *Ann. New York Acad. Sci. 752*:257-266 (1995)). Agents which promote angiogenesis, and particularly coronary artery angiogenesis, are therapeutically valuable to patients afflicted with vascular disease, and particularly heart disease.

Such agents promote the formation of collateral circulation and ameliorate the pathological effects of coronary artery occlusion.

Percutaneous transluminal coronary angioplasty (PTCA) is commonly used revascularization treatment for coronary artery occlusion and myocardial necrosis. However, coronary artery luminal narrowing (restenosis) after PTCA is an unfortunate complication which occurs in many patients (Rensing, B. J. *et al.*, *Circulation 88*:975-985 (1993)). There remains a need for therapeutic agents which can be used to prevent and treat restenosis.

## *Summary of the Invention*

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding the cerebellum and embryo specific protein (hereinafter "CESP") having the amino acid sequence shown in SEQ ID NO:2 or the amino acid sequence encoded by the cDNA clone deposited in a bacterial host as ATCC Deposit Number 97728 on September 23, 1996.

The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of CESP polypeptides or peptides by recombinant techniques.

The invention further provides an isolated CESP polypeptide having an amino acid sequence encoded by a polynucleotide described herein.

For a number of CESP-related disorders, it is believed that significantly higher or lower levels of CESP gene expression can be detected in certain tissues (e.g., heart, renal tubule, renal glomerulus, vascular endothelium, and aortic endothelium) or bodily fluids (e.g., blood, serum, plasma, urine, synovial fluid or spinal fluid, and amniotic fluid) taken from an individual having such a disorder, relative to a "standard" CESP gene expression level, i.e., the CESP expression

level in tissue or bodily fluids from an individual not having the CESP-related disorder. Thus, the invention provides a diagnostic method useful during diagnosis of a CESP-related disorder, which involves: (a) assaying CESP gene expression level in cells or body fluid of an individual; (b) comparing the CESP gene expression level with a standard CESP gene expression level, whereby an increase or decrease in the assayed CESP gene expression level compared to the standard expression level is indicative of a CESP-related disorder.

An additional aspect of the invention is related to a method for treating an individual in need of an increased level of CESP activity in the body comprising administering to such an individual a composition comprising a therapeutically effective amount of an isolated CESP polypeptide of the invention or an agonist thereof.

## *Brief Description of the Figures*

Figures 1A-1C show the nucleotide (SEQ ID NO:1) and deduced amino acid (SEQ ID NO:2) sequences of CESP. The protein has a leader sequence of about 21 amino acid residues (underlined) and a deduced molecular weight of about 38 kDa. The predicted amino acid sequence of the mature CESP protein is also shown.

Figure 2 shows the regions of similarity between the amino acid sequences of the CESP protein and a chicken gene for which the function is unknown (Genbank accession number D26311; SEQ ID NO:3).

Figure 3 shows an analysis of the CESP amino acid sequence. Alpha, beta, turn and coil regions; hydrophilicity and hydrophobicity; amphipathic regions; flexible regions; antigenic index and surface probability are shown. In the "Antigenic Index - Jameson-Wolf" graph, amino acid residues about 20 to about 86, about 92 to about 126, about 135 to about 157, about 169 to about 190, about 195 to about 219, about 234 to about 250, about 255 to about 274, and about 288 to about 336 in Figure 1 correspond to the shown highly antigenic regions of the

CESP receptor protein. These highly antigenic fragments in Figure 1 correspond to the following fragments, respectively, in SEQ ID NO:2: amino acid residues about -1 to about 65, about 71 to about 105, about 114 to about 136, about 148 to about 169, about 174 to about 198, about 213 to about 229, about 234 to about 253, and about 267 to about 315.

## *Detailed Description*

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding a CESP polypeptide having the amino acid sequence shown in SEQ ID NO:2. The CESP protein of the present invention shares sequence homology with a chicken gene for which the function is unknown (Figure 2; SEQ ID NO:3) (Genbank accession number D26311).

The amino acid sequence in SEQ ID NO:2 was deduced from the sequence of CESP cDNA clone HHFHG78. The nucleotide sequence shown in SEQ ID NO:1 was obtained by sequencing the HHFHG78 clone, which was deposited on September 23, 1996 at the American Type Culture Collection, 12301 Park Lawn Drive, Rockville, Maryland 20852, and given accession number 97728. In this clone, the CESP sequence is contained between EcoR I and Xho I sites in the polylinker of the pBluescript SK(-) plasmid (Stratagene, La Jolla, CA).

### *Nucleic Acid Molecules*

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the Model 373 from Applied Biosystems, Inc.), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain

some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence, compared to the actual sequence, will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

Using the information provided herein, such as the nucleotide sequence in SEQ ID NO:1, a nucleic acid molecule of the present invention encoding a CESP polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA as starting material. Illustrative of the invention, the nucleic acid molecule described in SEQ ID NO:1 was discovered in a cDNA library derived from human fetal heart. The determined nucleotide sequence of the CESP cDNA of SEQ ID NO:1 contains an open reading frame encoding a protein of 350 amino acid residues, and a deduced molecular weight of about 38 kDa. The CESP protein shown in SEQ ID NO:2 is about 58% identical and about 74% similar to an unknown chicken gene (Genbank accession number D26311) (Figure 2; SEQ ID NO:3).

The present invention also provides the mature form(s) of the CESP protein of the present invention. According to the signal hypothesis, proteins secreted by mammalian cells have a signal or secretory leader sequence which is cleaved from the mature protein once export of the growing protein chain across the rough endoplasmic reticulum has been initiated. Most mammalian cells and even insect cells cleave secreted proteins with the same specificity. However, in some cases, cleavage of a secreted protein is not entirely uniform, which results in two or more mature species on the protein. Further, it has long been known

that the cleavage specificity of a secreted protein is ultimately determined by the primary structure of the complete protein, that is, it is inherent in the amino acid sequence of the polypeptide. Therefore, the present invention provides a nucleotide sequence encoding the mature CESP polypeptides having the amino acid sequence encoded by the cDNA clone contained in the host identified as ATCC Deposit No. 97728 and as shown in SEQ ID NO:2. By the mature CESP protein having the amino acid sequence encoded by the cDNA clone contained in the host identified as ATCC Deposit 97728 is meant the mature form(s) of the CESP protein produced by expression in a mammalian cell (e.g., COS cells, as described below) of the complete open reading frame encoded by the human DNA sequence of the clone contained in the vector in the deposited host. As indicated below, the mature CESP protein having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728 may or may not differ from the predicted "mature" CESP protein shown in SEQ ID NO:2 (amino acids from about 1 to about 329 in SEQ ID NO:2), depending on the accuracy of the predicted cleavage site based on computer analysis.

Methods for predicting whether a protein has a secretory leader as well as the cleavage point for that leader sequence are available. For instance, the methods of McGeoch (*Virus Res. 3*:271-286 (1985)) and von Heinje (*Nucleic Acids Res. 14*:4683-4690 (1986)) can be used. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. von Heinje, *supra*. However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the predicted amino acid sequence of the complete CESP polypeptides of the present invention were analyzed by a computer program ("PSORT") (K. Nakai and M. Kanehisa, *Genomics 14*:897-911 (1992)), which is an expert system for predicting the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis by the PSORT program predicted the cleavage sites between amino acids -1 and 1 in

SEQ ID NO:2. Thereafter, the complete amino acid sequences were further analyzed by visual inspection, applying a simple form of the (-1,-3) rule of von Heinje. von Heinje, *supra*. Thus, the leader sequence for the CESP protein is predicted to consist of amino acid residues -21 to -1 in SEQ ID NO:2, while the predicted mature CESP protein consists of residues about 1 to about 329 in SEQ ID NO:2.

As one of ordinary skill would appreciate, due to the possibilities of sequencing errors discussed above, as well as the variability of cleavage sites for leaders in different known proteins, the full-length CESP polypeptide comprises about 350 amino acids, but may be anywhere in the range of 335 to 365 amino acids; and the predicted leader sequence of this protein is about 21 amino acids, but may be anywhere in the range of about 14 to about 50 amino acids.

As indicated, nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Isolated nucleic acid molecules of the present invention include DNA molecules comprising an open reading frame (ORF) shown in SEQ ID NO:1; DNA molecules comprising the coding sequence for the mature CESP protein

shown in SEQ ID NO:2 (last 329 amino acids); and DNA molecules which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the CESP protein. Of course, the genetic code is well known in the art. Thus, it would be routine for one skilled in the art to generate such degenerate variants.

In another aspect, the invention provides isolated nucleic acid molecules encoding the CESP polypeptide having an amino acid sequence as encoded by the cDNA clone contained in the plasmid deposited as ATCC Deposit No. 97728 on September 23, 1996. In further embodiments, this nucleic acid molecule will encode the mature polypeptide or the full-length polypeptide lacking the N-terminal methionine. The invention further provides an isolated nucleic acid molecule having the nucleotide sequence shown in SEQ ID NO:1 or the nucleic acid sequence of the CESP cDNA contained in the above-described deposited clone, or a nucleic acid molecule having a sequence complementary to one of the above sequences. Such isolated molecules, particularly DNA molecules, are useful as probes for gene mapping, by *in situ* hybridization with chromosomes, and for detecting expression of the CESP gene in human tissue, for instance, by Northern blot analysis.

The present invention is further directed to fragments of the isolated nucleic acid molecules described herein. By a fragment of an isolated nucleic acid molecule having the nucleotide sequence of the deposited cDNA or the nucleotide sequence shown in SEQ ID NO:1 is intended fragments at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length which are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments 50, 100, 150, 200, 250, 300, 350, 400, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, or 1100 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide sequence of the deposited cDNA or as shown in SEQ ID NO:1. By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or

more contiguous bases from the nucleotide sequence of the deposited cDNA or the nucleotide sequence as shown in SEQ ID NO:1.

Preferred nucleic acid fragments of the present invention also include nucleic acid molecules encoding epitope-bearing portions of the CESP protein. In particular, such nucleic acid fragments of the present invention include nucleic acid molecules encoding: a polypeptide comprising amino acid residues from about amino acid about -1 to about 65 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 71 to about 105 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 114 to about 136 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 148 to about 169 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 174 to about 198 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 213 to about 229 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 234 to about 253 in SEQ ID NO:2; and a polypeptide comprising amino acid residues from and about 267 to about 315 in SEQ ID NO:2.

In addition, the present inventors have identified the following cDNA clones related to extensive portions of the coding region of SEQ ID NO:1: HHFBI55Ra (SEQ ID NO:12); HHFDB95R (SEQ ID NO:13); HUSFC71R (SEQ ID NO:14); and HCE2S01R (SEQ ID NO:15). The present inventors have identified the following cDNA clone related to an extensive portion of the non-coding region of SEQ ID NO:1: HCEB157R (SEQ ID NO:16)

The following public ESTs, which relate to portions of the coding region of SEQ ID NO:1 have also been identified: GenBank Accession No. W61032 (SEQ ID NO:17); GenBank Accession No. AA349552 (SEQ ID NO:18); GenBank Accession No. R52311 (SEQ ID NO:19); GenBank Accession No. AA351624 (SEQ ID NO:20); GenBank Accession No. C05172 (SEQ ID NO:21); GenBank Accession No. T33818 (SEQ ID NO:22); GenBank Accession No. AA324686 (SEQ ID NO:23); GenBank Accession No. Z42237 (SEQ ID NO:24); GenBank Accession No. T30923 (SEQ ID NO:25); GenBank Accession No.

AA226979 (SEQ ID NO:26); GenBank Accession No. W45085 (SEQ ID NO:27); GenBank Accession No. T31076 (SEQ ID NO:28); GenBank Accession No. T08793 (SEQ ID NO:29); GenBank Accession No. R14945 (SEQ ID NO:30); GenBank Accession No. AA031480 (SEQ ID NO:31); GenBank Accession No. AA424460 (SEQ ID NO:32); GenBank Accession No. C05296 (SEQ ID NO:33); GenBank Accession No. R58671 (SEQ ID NO:34); GenBank Accession No. T18925 (SEQ ID NO:35);and GenBank Accession No. R57834 (SEQ ID NO:36).

In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a portion of the polynucleotide in a nucleic acid molecule of the invention described above, for instance, the cDNA clone contained in ATCC Deposit 97728. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 g/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably about 30, 40, 50, 60, or 70 nt of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below.

By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide (e.g., the deposited cDNA or the nucleotide sequence as shown in SEQ ID NO:1). Of course, a polynucleotide which hybridizes only to a poly A sequence (such as the 3′ terminal poly(A) tract of the CESP cDNA shown in SEQ ID NO:1), or to a complementary stretch of T (or U) resides, would not be included in a polynucleotide of the invention used to hybridize to a

portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

As indicated, nucleic acid molecules of the present invention which encode a CESP polypeptide may include, but are not limited to those encoding the amino acid sequence of the mature polypeptide, by itself; the coding sequence for the mature polypeptide and additional sequences, such as those encoding the about 21 amino acid leader or secretory sequence, such as a pre-, or pro- or prepro-protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5′ and 3′ sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing, including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; an additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities. Thus, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in Gentz *et al., Proc. Natl. Acad. Sci. USA 86*:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al., Cell 37*: 767 (1984). As discussed below, other such fusion proteins include the CESP protein fused to Fc at the N- or C-terminus.

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the CESP protein. Variants may occur naturally, such as a natural allelic

variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. *Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions, which may involve one or more nucleotides. The variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the CESP protein or portions thereof. Also especially preferred in this regard are conservative substitutions.

Further embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 95% identical, and more preferably at least 96%, 97%, 98% or 99% identical to (a) a nucleotide sequence encoding the polypeptide having the complete amino acid sequence in SEQ ID NO:2 (amino acid residues -21 to 329), including the predicted leader sequence; (b) a nucleotide sequence encoding the polypeptide having the complete amino acid sequence in SEQ ID NO:2 except for the N-terminal methionine (amino acid residues -20 to 329); (c) a nucleotide sequence encoding the polypeptide having the amino acid sequence at positions 1-329 in SEQ ID NO:2; (d) a nucleotide sequence encoding the polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; (e) a nucleotide sequence encoding the mature CESP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; or (f) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), or (e).

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a CESP polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the

reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the CESP polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5′ or 3′ terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 95%, 96%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in SEQ ID NO:1 or to the nucleotides sequence of the deposited cDNA clone can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711. Bestfit uses the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics 2*: 482-489 (1981), to find the best segment of homology between two sequences. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed.

The present application is directed to nucleic acid molecules at least 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in SEQ ID NO:1 or to the nucleic acid sequence of the deposited cDNA. This is because

even where a particular nucleic acid molecule does not encode a polypeptide having CESP activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having CESP activity include, *inter alia*, (1) isolating the CESP gene or allelic variants thereof in a cDNA library; (2) *in situ* hybridization (e.g., "FISH") to metaphase chromosomal spreads to provide precise chromosomal location of the CESP gene, as described in Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York (1988); and (3) Northern Blot analysis for detecting CESP mRNA expression in specific tissues.

Preferred, however, are nucleic acid molecules having sequences at least 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in SEQ ID NO:1 or to the nucleic acid sequence of the deposited cDNA which do, in fact, encode a polypeptide having CESP activity. By "a polypeptide having CESP activity" is intended polypeptides exhibiting activity in a particular biological assay. For example, protein activity can be measured using the morphometrically quantitative *in vitro* assay for angiogenesis as described by Sueishi *et al.* (*Japanese Circulation J. 56*:192-198 (1992). This assay utilizes a model of angiogenesis in a culture system using type I collagen gel s a reconstructed subendothelial matrix. The length of capillary-like tubular structures are measured morphometrically using an image analyzer. Briefly, this assay involves isolating and culturing capillary endothelial cells (for example, from bovine adrenal cortex or another suitable source), administering a candidate protein to the cell culture, and measuring morphometrically the total length of tubular structures using phase-contrast microscopic photography.

Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in SEQ ID NO:1 or to the nucleic acid sequence

of the deposited cDNA will encode a polypeptide having CESP protein activity. In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having CESP protein activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid).

For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science 247:*1306-1310 (1990), wherein the authors indicate that proteins are surprisingly tolerant of amino acid substitutions.

***Vectors and Host Cells***

The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of CESP polypeptides or fragments thereof by recombinant techniques.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli lac, trp* and *tac* promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression

constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will preferably include a translation initiating at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture and tetracycline or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate heterologous hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al., Basic Methods In Molecular Biology* (1986).

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of additional amino acids,

particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to solubilize proteins. For example, EP-A-O 464 533 (Canadian counterpart 2,045,869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is thoroughly advantageous for use in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified in the advantageous manner described. This is the case when Fc portion proves to be a hindrance to use in therapy and diagnosis, for example when the fusion protein is to be used as antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. See, D. Bennett *et al.*, *Journal of Molecular Recognition*, Vol. 8:52-58 (1995) and K. Johanson *et al.*, *The Journal of Biological Chemistry*, Vol. 270, No. 16:9459-9471 (1995).

The CESP protein can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography

("HPLC") is employed for purification. Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

***CESP Polypeptides and Fragments***

The invention further provides an isolated CESP polypeptide having the amino acid sequence encoded by the deposited cDNA, or the amino acid sequence in SEQ ID NO:2, or a peptide or polypeptide comprising a portion of the above polypeptides.

It will be recognized in the art that some amino acid sequences of the CESP polypeptide can be varied without significant effect of the structure or function of the protein. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity.

Thus, the invention further includes variations of the CESP polypeptide which show substantial CESP polypeptide activity or which include regions of CESP protein such as the protein portions discussed below. Such mutants include deletions, insertions, inversions, repeats, and type substitutions. As indicated above, guidance concerning which amino acid changes are likely to be phenotypically silent can be found in Bowie, J.U., *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science 247:*1306-1310 (1990).

Thus, the fragment, derivative or analog of the polypeptide of SEQ ID NO:2, or that encoded by the deposited cDNA, may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as an IgG Fc fusion region peptide or leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Of particular interest are substitutions of charged amino acids with another charged amino acid and with neutral or negatively charged amino acids. The latter results in proteins with reduced positive charge to improve the characteristics of the CESP protein. The prevention of aggregation is highly desirable. Aggregation of proteins not only results in a loss of activity but can also be problematic when preparing pharmaceutical formulations, because they can be immunogenic (Pinckard *et al., Clin. Exp. Immunol. 2*:331-340 (1967); Robbins *et al., Diabetes 36*:838-845 (1987); and Cleland *et al. Crit. Rev. Therapeutic Drug Carrier Systems 10*:307-377 (1993)).

The replacement of amino acids can also change the selectivity of binding to cell surface receptors. Ostade *et al., Nature 361*:266-268 (1993), describes certain mutations resulting in selective binding of TNF-α to only one of the two known types of TNF receptors. Thus, the CESP protein of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation.

As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein (see Table 1).

Amino acids in the CESP protein of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science 244*:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro*, or *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol. 224*:899-904 (1992) and de Vos *et al. Science 255*:306-312 (1992)).

TABLE 1. Conservative Amino Acid Substitutions.

| | |
|---|---|
| Aromatic | Phenylalanine<br>Tryptophan<br>Tyrosine |
| Hydrophobic | Leucine<br>Isoleucine<br>Valine |
| Polar | Glutamine<br>Asparagine |
| Basic | Arginine<br>Lysine<br>Histidine |
| Acidic | Aspartic Acid<br>Glutamic Acid |
| Small | Alanine<br>Serine<br>Threonine<br>Methionine<br>Glycine |

Of course, the number of amino acid substitutions a skilled artisan would make depends on many factors, including those described above. Generally speaking, the number of amino acid substitutions for any given CESP polypeptide will not be more than 50, 40, 30, 20, 10, 5, or 3.

Amino acids in the CESP protein of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science 244*:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity, such as *in vitro* proliferative activity.

The polypeptides of the present invention are preferably provided in an isolated form. By "isolated polypeptide" is intended a polypeptide removed from its native environment. Thus, a polypeptide produced or contained in a recombinant host cell is considered "isolated" for the purposes of the present invention. Also intended as "isolated " is a polypeptide that has been purified, partially or substantially, from a recombinant host or from a native source. For example, a recombinantly produced version of the CESP polypeptide can be substantially purified by the one-step method described in Smith and Johnson, *Gene 67*:31-40 (1988).

The polypeptides of the present invention include the complete polypeptide encoded by the deposited cDNA; the mature polypeptide encoded by the deposited cDNA; amino acid residues -21 to 329 of SEQ ID NO:2; amino acid residues -20 to 329 of SEQ ID NO:2; and amino acid residues 1 to 329 in SEQ ID NO:2, as well as polypeptides which are at least 95% identical, and more preferably at least 96%, 97%, 98% or 99% identical to the polypeptide encoded by the deposited cDNA, and to the polypeptides of SEQ ID NO:2, and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a CESP polypeptide is

intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of the CESP polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence shown in SEQ ID NO:2 or to the amino acid sequence encoded by deposited cDNA clone can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

The polypeptide of the present invention could be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention. The

epitope of this polypeptide portion is an immunogenic or antigenic epitope of a polypeptide described herein. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen *et al., Proc. Natl. Acad. Sci. USA 81*:3998-4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in the art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Learner, R.A., Antibodies that react with predetermined sites on proteins, *Science 219:*660-666 (1983). Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals.

Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance, Wilson *et al., Cell 37*:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about at least about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention.

Non-limiting examples of antigenic polypeptides or peptides that can be used to generate CESP-specific antibodies include: a polypeptide comprising amino acid residues from about amino acid about -1 to about 65 in SEQ ID NO:2;

a polypeptide comprising amino acid residues from about 71 to about 105 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 114 to about 136 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 148 to about 169 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 174 to about 198 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 213 to about 229 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 234 to about 253 in SEQ ID NO:2; and a polypeptide comprising amino acid residues from and about 267 to about 315 in SEQ ID NO:2.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means (Houghten, R. A., General method for the rapid solid-phase synthesis of large numbers of peptides: Specificity of antigen-antibody interaction at the level of individual amino acids, *Proc. Natl. Acad. Sci. USA 82*:5131-5135 (1985)). This "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten *et al.* (1986).

As one of skill in the art will appreciate, CESP polypeptides of the present invention and the epitope-bearing fragments thereof described above can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life *in vivo.* This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EPA 394,827; Traunecker *et al., Nature 331*:84- 86 (1988)). Fusion proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than the monomeric CESP protein or protein fragment alone (Fountoulakis *et al., J Biochem. 270*:3958-3964 (1995)).

*Diagnostic and Prognostic Applications of CESP*

It is believed that certain tissues in mammals with a CESP-related disorder express significantly enhanced or diminished levels of the CESP protein and mRNA encoding the CESP protein when compared to a corresponding "standard" mammal, i.e., a mammal of the same species not having the disorder. Further, it is believed that enhanced or diminished levels of the CESP protein can be detected in certain body fluids (e.g., blood, sera, plasma, urine, and spinal fluid) from mammals with the disorder when compared to sera from mammals of the same species not having the disorder. Thus, the invention provides a diagnostic method useful during diagnosis, which involves assaying the expression level of the gene encoding the CESP protein in mammalian cells or body fluid and comparing the gene expression level with a standard CESP gene expression level, whereby an increase in the gene expression level over the standard is indicative of certain disorders.

CESP related disorders include but are not limited to coronary restenosis following coronary revascularization, coronary artery thrombus or occlusion, myocardial infarction, atrial and/or ventricular arrhythmias, heart block, hereditary medial "necrosis" of small coronary and pulmonary arteries, focal fibromuscular dysplasia of small coronary arteries, cardiomyopathy, arrhythmogenic right ventricular dysplasia, and sudden death.

Where a diagnosis has already been made according to conventional methods, the present invention is useful as a prognostic indicator, whereby patients exhibiting enhanced or decreased CESP gene expression will experience a worse clinical outcome relative to patients expressing the gene at a lower level. Further, CESP is detected in amniotic cells. It is believed that CESP can serve as a marker for fetal genetic defects. Such fetal genetic defects include developmental cardiac defects.

By "assaying the expression level of the gene encoding the CESP protein" is intended qualitatively or quantitatively measuring or estimating the level of the

CESP protein or the level of the mRNA encoding the CESP protein in a first biological sample either directly (e.g., by determining or estimating absolute protein level or mRNA level) or relatively (e.g., by comparing to the CESP protein level or mRNA level in a second biological sample).

Preferably, the CESP protein level or mRNA level in the first biological sample is measured or estimated and compared to a standard CESP protein level or mRNA level, the standard being taken from a second biological sample obtained from an individual not having the disorder. As will be appreciated in the art, once a standard CESP protein level or mRNA level is known, it can be used repeatedly as a standard for comparison.

By "biological sample" is intended any biological sample obtained from an individual, cell line, tissue culture, or other source which contains CESP protein or mRNA. Biological samples include mammalian body fluids (such as blood, sera, plasma, urine, synovial fluid, spinal fluid, and amniotic fluid containing amniotic cells) which contain secreted mature CESP protein, heart, renal glomerulus, and renal tubule.

The present invention is useful for detecting CESP-related disorders in mammals. Preferred mammals include monkeys, apes, cats, dogs, cows, pigs, horses, rabbits and humans. Particularly preferred are humans.

Total cellular RNA can be isolated from a biological sample using the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski and Sacchi, *Anal. Biochem. 162:*156-159 (1987). Levels of mRNA encoding the CESP protein are then assayed using any appropriate method. These include Northern blot analysis Harada *et al., Cell 63*:303-312 (1990)), S1 nuclease mapping (Fujita *et al., Cell 49*:357-367 (1987)), the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR) (Makino *et al., Technique 2*:295-301 (1990)), and reverse transcription in combination with the ligase chain reaction (RT-LCR).

Assaying CESP protein levels in a biological sample can occur using antibody-based techniques. For example, CESP protein expression in tissues can

be studied with classical immunohistological methods (Jalkanen, M., *et al., J. Cell. Biol. 101:*976-985 (1985); Jalkanen, M., *et al., J. Cell . Biol. 105:*3087-3096 (1987)).

Other antibody-based methods useful for detecting CESP protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable labels are known in the art and include enzyme labels, such as glucose oxidase, and radioisotopes, such as iodine ($^{125}I$, $^{121}I$), carbon ($^{14}C$), sulphur ($^{35}S$), tritium ($^{3}H$), indium ($^{112}In$), and technetium ($^{99m}Tc$), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

*CESP Protein Therapy*

It is believed that HSF plays a role in a wide variety of physiologic and pathologic processes. Accordingly, the CESP protein has application to any physiologic or pathologic disease condition in which abnormal activity of the CESP system is implicated and has pathological or physiological consequences.

Physiological processes in which CESP is believed to be involved include the regulation of collateral circulation (particularly in the heart) regulation of coronary artery restenosis following a revascularization procedure, regulation of apoptosis in myocytes, the modulation of myocyte development in the developing heart, regulation of circulating blood volume, regulation of vascular tone, regulation of blood pressure and cardiac output, diuresis, natriuresis, facilitation of transudation of plasma water to the interstitium, and inhibition of the release or action of hormones such as aldosterone, angiotensin II, endothelins, renin, and vasopressin.

It is also believed that CESP plays a role as a growth modulator in the developing heart. Moreover, it is believed that CESP protects adult myocardial cells from damage during myocardial ischemia. Further, it is believed that CESP enhances revascularization of cardiac muscle following revascularization therapy

(e.g., coronary artery bypass surgery; percutaneous transluminal coronary angioplasty; or administration of an anticoagulant such as heparin, hirudin, urokinase, streptokinase, or tissue plasminogen activator) and prevents or inhibits restenosis of coronary arteries following revascularization therapy. Accordingly, when CESP is administered to a patient receiving revascularization therapy, CESP enhances revascularization of cardiac muscle and prevents or inhibits restenosis of coronary arteries.

It is also believed that CESP facilitates angiogenesis (i.e., the formation of new vascular tissue). Accordingly, administration of CESP to patients afflicted by circulatory illnesses facilitates angiogenesis. Circulatory illness for which CESP treatment is beneficial include atherosclerotic heart disease, coronary artery constriction, coronary artery blockage (either partial or full), myocardial infarction, venous thrombosis, and Reynaud's syndrome.

The present invention is useful for treating or preventing CESP-related disorders in mammals. Preferred mammals include monkeys, apes, cats, dogs, cows, pigs, horses, rabbits and humans. Particularly preferred are humans.

*Modes of administration*

It will be appreciated that conditions caused by a decrease in the standard or normal level of CESP activity in an individual, can be treated by administration of CESP protein. Thus, the invention further provides a method of treating an individual in need of an increased level of CESP activity comprising administering to such an individual a pharmaceutical composition comprising an effective amount of an isolated CESP polypeptide of the invention, particularly a mature form of the CESP, effective to increase the CESP activity level in such an individual.

As a general proposition, the total pharmaceutically effective amount of CESP polypeptide administered parenterally per dose will be in the range of about 1 μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above,

this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the CESP polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed.

Pharmaceutical compositions containing the CESP protein of the invention may be administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, drops or transdermal patch), bucally, or as an oral or nasal spray. By "pharmaceutically acceptable carrier" is meant a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

*Chromosome Assays*

The nucleic acid molecules of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

In certain preferred embodiments in this regard, the cDNA herein disclosed is used to clone genomic DNA of a CESP protein gene. This can be accomplished using a variety of well known techniques and libraries, which generally are available commercially. The genomic DNA then is used for *in situ* chromosome mapping using well known techniques for this purpose.

In addition, in some cases, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3′ untranslated region of the gene is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes.

Fluorescence *in situ* hybridization ("FISH") of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with probes from the cDNA as short as 50 or 60 bp. For a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual Of Basic Techniques*, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, *Mendelian Inheritance In Man*, available on-line through Johns Hopkins University, Welch Medical Library. The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

## *Examples*

### *Example 1: Expression and Purification of CESP in E. coli*

The bacterial expression vector pQE60 is used for bacterial expression in this example. (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). pQE60 encodes ampicillin antibiotic resistance ("Amp$^r$") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), six codons encoding histidine residues that allow affinity purification using nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin sold by QIAGEN, Inc., *supra*, and suitable single restriction enzyme cleavage sites. These elements are arranged such that an inserted DNA fragment encoding a polypeptide expresses that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the carboxyl terminus of that polypeptide.

The DNA sequence encoding the desired portion of the CESP protein lacking the hydrophobic leader sequence is amplified from the deposited cDNA clone using PCR oligonucleotide primers which anneal to the amino terminal sequences of the desired portion of the CESP protein and to sequences in the deposited construct 3' to the cDNA coding sequence. Additional nucleotides containing restriction sites to facilitate cloning in the pQE60 vector are added to the 5' and 3' sequences, respectively.

For cloning the mature protein, the 5' primer has the sequence 5'-GGGA-GGATCCGCGCCCGCTCCGACGGCG-3' (SEQ ID NO:4), containing the underlined BamH I restriction site followed by 20 nucleotides corresponding to nucleotides 134-153 of the CESP cDNA sequence set out in SEQ ID NO:1. One of ordinary skill in the art would appreciate, of course, that the point in the protein coding sequence where the 5' primer begins may be varied to amplify a DNA segment encoding any desired portion of the complete protein shorter or longer than the mature form. The 3' primer has the sequence 5'-GCCTCTAGATTAAATCTCTTCCCCTCCCAGCAGT-3' (SEQ ID NO:5),

containing the underlined Xba I restriction site followed by 24 nucleotides complementary to nucleotides 1101 to 1124 of the CESP DNA sequence set out in SEQ ID NO:1, with the coding sequence aligned with the restriction site so as to maintain its reading frame with that of the six His codons in the pQE60 vector.

The amplified CESP DNA fragment and the vector pQE60 are digested with BamH I and Xba I restriction enzymes and the digested DNAs are then ligated together. Insertion of the CESP DNA into the restricted pQE60 vector places the CESP protein coding region downstream from the IPTG-inducible promoter and in-frame with an initiating AUG and the six histidine codons.

The ligation mixture is transformed into competent *E. coli* cells using standard procedures such as those described in Sambrook *et al.*, *Molecular Cloning: a Laboratory Manual, 2nd Ed.;* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989). *E. coli* strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses the lac repressor and confers kanamycin resistance ("Kan$^r$"), is used in carrying out the illustrative example described herein. This strain, which is only one of many that are suitable for expressing CESP protein, is available commercially from QIAGEN, Inc., *supra.* Transformants are identified by their ability to grow on LB plates in the presence of ampicillin and kanamycin. Plasmid DNA is isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs are grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 µg/ml) and kanamycin (25 µg/ml). The O/N culture is used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells are grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. Isopropyl-b-D-thiogalactopyranoside ("IPTG") is then added to a final concentration of 1 mM to induce transcription from the lac repressor sensitive promoter, by inactivating the lacI repressor. Cells subsequently are incubated further for 3 to 4 hours. Cells then are harvested by centrifugation.

The cells are then stirred for 3-4 hours at 4°C in 6M guanidine-HCl, pH8. The cell debris is removed by centrifugation, and the supernatant containing the CESP protein is loaded onto a nickel-nitrilo-tri-acetic acid ("NiNTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the NI-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist, 1995, QIAGEN, Inc., *supra).* Briefly the supernatant is loaded onto the column in 6 M guanidine-HCl, pH8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH8, then washed with 10 volumes of 6 M guanidine-HCl pH6, and finally the CESP protein is eluted with 6 M guanidine-HCl, pH5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins can be eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

*Example 2: Cloning and Expression of CESP protein in a Baculovirus Expression System*

The cDNA sequence encoding the full length CESP protein in the deposited clone was amplified using PCR oligonucleotide primers corresponding to the 5′ and 3′ sequences of the gene:

The 5′ primer had the sequence 5′-TGCCGCGGATCCGCCATCATG CAGCGGCTTGGGGCCAC-3′ (SEQ ID NO:6), containing the underlined BamH I restriction enzyme site followed by 20 nucleotides corresponding to nucleotides 73-92 of the CESP protein coding sequence set out in SEQ ID NO:1. Inserted

into an expression vector, as described below, the 5′ end of the amplified fragment encoding CESP provided an efficient signal peptide. An efficient signal for initiation of translation in eukaryotic cells, as described by Kozak, M., *J. Mol. Biol. 196*:947-950 (1987) was appropriately located in the vector portion of the construct.

The 3′ primer had the sequence 5′GCACAGGTACCCACAGCCTGGTC-CAGATCTAAATCTCTTCCCCTCCCAG 3′ (SEQ ID NO:7), containing the underlined Asp718 restriction site followed by 42 nucleotides complementary to nucleotides 1105-1145 of the CESP cDNA sequence set out in SEQ ID NO:1.

The amplified fragment was isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then was digested with BamH I and Asp718 and again was purified on a 1% agarose gel. This fragment is designated herein F2.

The vector pA2 was used to express the CESP protein in the baculovirus expression system, using standard methods, as described in Summers *et al.*, A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures, Texas Agricultural Experimental Station Bulletin No. 1555 (1987). This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For an easy selection of recombinant virus the beta-galactosidase gene from *E. coli* is inserted in the same orientation as the polyhedrin promoter and is followed by the polyadenylation signal of the polyhedrin gene. The polyhedrin sequences are flanked at both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that expresses the cloned polynucleotide.

The pA2 expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For an easy selection of

recombinant virus the beta-galactosidase gene from *E. coli* is inserted in the same orientation as the polyhedrin promoter and is followed by the polyadenylation signal of the polyhedrin gene. The polyhedrin sequences are flanked at both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate viable virus that express the cloned polynucleotide.

Many other baculovirus vectors could be used in place of pA2, such as pA2-GP (which contains the AcMNPV gp 67 signal peptide), pAc373, pVL941 and pAcIM1 provided, as those of skill readily will appreciate, that construction provides appropriately located signals for transcription, translation, trafficking and the like, such as an in-frame AUG and a signal peptide, as required. Such vectors are described in Luckow *et al.*, *Virology 170*: 31-39, among others.

The pA2 plasmid was digested with the restriction enzymes BamH I and Asp718. The DNA was then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated herein "V".

Fragment F2 and the dephosphorylated plasmid V2 were ligated together with T4 DNA ligase. *E. coli* HB101 cells were transformed with ligation mix and spread on culture plates. Bacteria were identified that contained the plasmid with the human CESP gene by digesting DNA from individual colonies using BamH I and Asp718 and then analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment was confirmed by DNA sequencing. This plasmid is designated herein pBacCESP.

5 μg of the plasmid pBacCESP were co-transfected with 1.0 μg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA.), using the lipofection method described by Felgner *et al.*, Proc. Natl. Acad. Sci. USA 84: 7413-7417 (1987). 1μg of BaculoGold™ virus DNA and 5 μg of the plasmid pBacCESP were mixed in a sterile well of a microtiter plate containing 50 μl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μl Lipofectin plus 90 μl Grace's medium are added, mixed and incubated for 15 minutes at room

temperature. Then the transfection mixture was added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate was rocked back and forth to mix the newly added solution. The plate was then incubated for 5 hours at 27°C. After 5 hours, the transfection solution was removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum was added. The plate was put back into an incubator and cultivation was continued at 27°C for four days.

After four days, the supernatant was collected and a plaque assay was performed, as described by Summers and Smith, cited above. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) was used to allow easy identification and isolation of gal-expressing clones, which produced blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10).

Four days after serial dilution, the virus was added to the cells. After appropriate incubation, blue stained plaques were picked with the tip of an Eppendorf pipette. The agar containing the recombinant viruses was then resuspended in an Eppendorf tube containing 200 µl of Grace's medium. The agar was removed by a brief centrifugation and the supernatant containing the recombinant baculovirus was used to infect Sf9 cells seeded in 35 mm dishes. Four days later, the supernatants of these culture dishes were harvested and then they were stored at 4°C. A clone containing properly inserted hESSB I, II and III was identified by DNA analysis including restriction mapping and sequencing. This is designated herein as V-CESP.

Sf9 cells were grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells were infected with the recombinant baculovirus V-CESP at a multiplicity of infection ("MOI") of about 2 (about 1 to about 3). Six hours later, the medium was removed and was replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Gaithersburg). 42 hours later, 5 µCi of $^{35}$S-methionine and 5 µCi $^{35}$S-cysteine

(available from Amersham) were added. The cells were further incubated for 16 hours and then they were harvested by centrifugation, lysed and the labeled proteins were visualized by SDS-PAGE and autoradiography.

*Example 3: Cloning and Expression in Mammalian Cells*

A typical mammalian expression vector contains the promoter element, which mediates the initiation of transcription of mRNA, the protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription can be achieved with the early and late promoters from SV40, the long terminal repeats (LTRS) from retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter). Suitable expression vectors for use in practicing the present invention include, for example, vectors such as PSVL and PMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146) and pBC12MI (ATCC 67109). Mammalian host cells that could be used include human HeLa 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV 1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the gene can be expressed in stable cell lines that contain the gene integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, or hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful to develop cell lines that carry several hundred or even several thousand copies of the gene of interest. Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy *et al., Biochem J. 227*:277-279 (1991); Bebbington *et al.,*

*Bio/Technology 10:*169-175 (1992)). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

The expression vectors pC1 and pC4 contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen *et al., Molecular and Cellular Biology*, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart *et al., Cell 41*:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamH I, Xba I and Asp 718, facilitate the cloning of the gene of interest. The vectors contain in addition the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene.

***Example 3(a): Cloning and Expression in COS Cells***

The expression plasmid, pCESP HA, is made by cloning a cDNA encoding CESP into the expression vector pcDNAI/Amp or pcDNA3 (which can be obtained from Invitrogen, Inc.).

The expression vector pcDNA3 contains: (1) an *E. coli* origin of replication effective for propagation in *E. coli* and other prokaryotic cells; (2) an ampicillin resistance gene for selection of plasmid-containing prokaryotic cells; (3) an SV40 origin of replication for propagation in eukaryotic cells; (4) a CMV promoter, a polylinker, an SV40 intron; (5) several codons encoding a hemagglutinin fragment (i.e., an "HA" tag to facilitate purification) followed by a termination codon and polyadenylation signal arranged so that a cDNA can be conveniently placed under expression control of the CMV promoter and operably linked to the SV40 intron and the polyadenylation signal by means of restriction sites in the polylinker. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein described by Wilson *et al., Cell 37:*767 (1984). The fusion of the HA tag to the target protein allows easy detection and recovery

of the recombinant protein with an antibody that recognizes the HA epitope. pcDNA3 contains, in addition, the selectable neomycin marker.

A DNA fragment encoding the CESP is cloned into the polylinker region of the vector so that recombinant protein expression is directed by the CMV promoter. The plasmid construction strategy is as follows. The CESP cDNA of the deposited clone is amplified using primers that contain convenient restriction sites, much as described above for construction of vectors for expression of CESP in *E. coli.* Suitable primers include the following, which are used in this example. The 5' primer has the sequence 5′-TGCCGCGGATCCGCCATCATG CAGCGGCTTGGGGCCAC-3′ (SEQ ID NO:6), containing the underlined BamH I restriction enzyme site, a Kozak sequence, and an AUG start codon followed by 20 nucleotides (nucleotides 73-92) of the CESP protein coding sequence in SEQ ID NO:1. If no HA tag is used, the 3' primer has the sequence 5'-GTCTCTAGACAGATCTAAATCTCTTCCCCTCCCAG-3' (SEQ ID NO:8), containing the underlined Xba I site and 26 nucleotides complementary to nucleotides 1105-1130 of the CESP cDNA sequence set out in SEQ ID NO:1. If an HA tag is used, the 3' primer has the sequence 5'-GTCTCTAGACAGA-TCTAAGCGTAGTCTGGGACGTCGTATGGGTAAATCTCTTCCCCTCCC-AGCAG-3' (SEQ ID NO:9), containing the underlined Xba I site and 23 nucleotides complementary to nucleotides 1102-1124 of the CESP cDNA sequence set out in SEQ ID NO:1

The PCR amplified DNA fragment and the vector, pcDNA3, are digested with Xba I restriction enzyme and then ligated. The ligation mixture is transformed into *E. coli* strain SURE (available from Stratagene Cloning Systems, 11099 North Torrey Pines Road, La Jolla, CA 92037), and the transformed culture is plated on ampicillin media plates which then are incubated to allow growth of ampicillin resistant colonies. Plasmid DNA is isolated from resistant colonies and examined by restriction analysis or other means for the presence of the CESP-encoding fragment.

For expression of recombinant CESP, COS cells are transfected with an expression vector, as described above, using DEAE-DEXTRAN, as described, for instance, in Sambrook *et al., Molecular Cloning: a Laboratory Manual,* Cold Spring Laboratory Press, Cold Spring Harbor, New York (1989). Cells are incubated under conditions for expression of CESP by the vector.

Expression of the CESP-HA fusion protein is detected by radiolabeling and immunoprecipitation, using methods described in, for example Harlow *et al., Antibodies: A Laboratory Manual, 2nd Ed.;* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988). To this end, two days after transfection, the cells are labeled by incubation in media containing $^{35}$S-cysteine for 8 hours. The cells and the media are collected, and the cells are washed and lysed with detergent-containing RIPA buffer: 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM TRIS, pH 7.5, as described by Wilson *et al.* cited above. Proteins are precipitated from the cell lysate and from the culture media using an HA-specific monoclonal antibody. The precipitated proteins then are analyzed by SDS-PAGE and autoradiography. An expression product of the expected size is seen in the cell lysate, which is not seen in negative controls.

***Example 3(b): Cloning and Expression in CHO Cells***

The vector pC4 was used for the expression of the CESP protein. Plasmid pC4 is a derivative of the plasmid pSV2-dhfr (ATCC Accession No. 37146). The plasmid contains the mouse DHFR gene under control of the SV40 early promoter. Chinese hamster ovary- or other cells lacking dihydrofolate activity that are transfected with these plasmids can be selected by growing the cells in a selective medium (alpha minus MEM, Life Technologies) supplemented with the chemotherapeutic agent methotrexate. The amplification of the DHFR genes in cells resistant to methotrexate (MTX) has been well documented (see, e.g., Alt, F. W., Kellems, R. M., Bertino, J. R., and Schimke, R. T., 1978, *J Biol. Chem. 253*:1357-1370, Hamlin, J. L. and Ma, C. 1990, *Biochem. et Biophys. Acta,*

*1097*:107-143, Page, M. J. and Sydenham, M.A. 1991, *Biotechnology 9*:64-68). Cells grown in increasing concentrations of MTX develop resistance to the drug by overproducing the target enzyme, DHFR, as a result of amplification of the DHFR gene. If a second gene is linked to the DHFR gene, it is usually co-amplified and over-expressed. It is known in the art that this approach may be used to develop cell lines carrying more than 1,000 copies of the amplified gene(s). Subsequently, when the methotrexate is withdrawn, cell lines are obtained which contain the amplified gene integrated into one or more chromosome(s) of the host cell.

Plasmid pC4 contains for expressing the gene of interest the strong promoter of the long terminal repeat (LTR) of the Rous Sarcoma Virus (Cullen, *et al., Molecular and Cellular Biology*, March 1985:438-447) plus a fragment isolated from the enhancer of the immediate early gene of human cytomegalovirus (CMV) (Boshart *et al., Cell 41*:521-530 (1985)). Downstream of the promoter are BamH I, Xba I, and Asp 718 restriction enzyme cleavage sites that allow integration of the genes. Behind these cloning sites the plasmid contains the 3' intron and polyadenylation site of the rat preproinsulin gene. Other high efficiency promoters can also be used for the expression, e.g., the human β-actin promoter, the SV40 early or late promoters or the long terminal repeats from other retroviruses, e.g., HIV and HTLVI. Clontech's Tet-Off and Tet-On gene expression systems and similar systems can be used to express the CESP protein in a regulated way in mammalian cells (Gossen, M., & Bujard, H. 1992, *Proc. Natl. Acad. Sci. USA 89*: 5547-5551). For the polyadenylation of the mRNA other signals, e.g., from the human growth hormone or globin genes can be used as well. Stable cell lines carrying a gene of interest integrated into the chromosomes can also be selected upon co-transfection with a selectable marker such as gpt, G418 or hygromycin. It is advantageous to use more than one selectable marker in the beginning, e.g., G418 plus methotrexate.

The plasmid pC4 was digested with the restriction enzyme BamH I and then dephosphorylated using calf intestinal phosphatase by procedures known in the art. The vector was then isolated from a 1% agarose gel.

The DNA sequence encoding the complete CESP protein, including its leader sequence, was amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene. The 5' primer had the sequence 5'-GC TGCCGCGGATCCGCCACCATGCAGCGGCTTGGGGCCACC 3' (SEQ ID NO:10 containing the underlined BamH I restriction enzyme site, an efficient signal for initiation of translation in eukaryotes, as described by Kozak, M., *J. Mol. Biol. 196*:947-950 (1987), and followed by 21 nucleotides corresponding to nucleotides 73-93 of the CESP protein coding sequence set out in SEQ ID NO:1.

The 3' primer had the sequence 5'-CACACGCGGATCCAGATCTAAA TCTCTTCCCCTC-3' (SEQ ID NO:11) containing the underlined BamH I restriction site followed by 24 nucleotides (nucleotides 1109-1132) complementary to the CESP protein coding sequence set out in SEQ ID NO:1, including the stop codon.

The amplified fragment was digested with the restriction enzyme BamH I and then purified again on a 1% agarose gel. The isolated fragment and the dephosphorylated vector were then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells were then transformed and bacteria were identified that contain the fragment inserted into plasmid pC4 using restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene were used for transfection. 5 μg of the expression plasmid pC4 were cotransfected with 0.5 μg of the plasmid pSV2-neo using lipofectin (Felgner *et al., supra).* The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells were seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells were trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones

were trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate were then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μM, 2 μM, 5 μM, 10 mM, 20 mM). The same procedure was repeated until clones were obtained which grow at a concentration of 100 - 200 μM. Expression of the desired gene product was analyzed by SDS-PAGE and Western blot or by reverse phase HPLC analysis.

***Example 4: Tissue distribution of CESP protein expression***

Northern blot analysis was carried out to examine CESP gene expression in human tissues, using the methods described by Sambrook *et al.*, cited above. A cDNA probe containing the entire nucleotide sequence of the CESP protein (SEQ ID NO:1) was labeled with $^{32}P$ using the *redi*prime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labelling, the probe was purified using a CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labelled probe was then used to examine various human tissues for CESP mRNA.

Multiple Tissue Northern (MTN) blots from human heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, and peripheral blood leukocytes were obtained from Clontech and were examined with labelled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots were mounted and exposed to film at -70°C overnight, and films developed according to standard procedures. An abundant 2.6 kilobase transcript was detected in heart and brain. A weaker 2.6 kilobase signal was detected in endothelial cells, amniotic cells, smooth muscle, HSC172 cells and osteoblastoma cells.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of all publications (including patents, patent applications, journal articles, laboratory manuals, books, or other documents) cited herein are hereby incorporated by reference.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HUMAN GENOME SCIENCES, INC.
9410 KEY WEST AVENUE
ROCKVILLE, MD 20850
UNITED STATES OF AMERICA

APPLICANT/INVENTOR: SOPPET, DANIEL R.
RUBEN, STEVEN M.

(ii) TITLE OF INVENTION: CEREBELLUM AND EMBRYO SPECIFIC PROTEIN

(iii) NUMBER OF SEQUENCES: 36

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
(B) STREET: 1100 NEW YORK AVENUE, SUITE 600
(C) CITY: WASHINGTON
(D) STATE: DC
(E) COUNTRY: US
(F) ZIP: 20005-3934

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: To be assigned
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/033,870
(B) FILING DATE: 20-DEC-1996
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: STEFFE, ERIC K.
(B) REGISTRATION NUMBER: 36,688
(C) REFERENCE/DOCKET NUMBER: 1488.061PC01

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (202) 371-2600
(B) TELEFAX: (202) 371-2540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2490 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 73..1122

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 73..133

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 136..1122

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AGAGGATCCG GGTTCGGTTG CTCTGGCGAG GGCTCCAGCA TCACAGGCGG CGGCTGCGGG       60

CGCAGAGCGG AG ATG CAG CGG CTT GGG GCC ACC CTG CTG TGC CTG CTG           108
              Met Gln Arg Leu Gly Ala Thr Leu Leu Cys Leu Leu
              -21 -20                 -15                 -10

CTG GCG GCG GCG GTC CCC ACG GCC CCC GCG CCC GCT CCG ACG GCG ACC        156
Leu Ala Ala Ala Val Pro Thr Ala Pro Ala Pro Ala Pro Thr Ala Thr
                  -5                   1                   5

TCG GCT CCA GTC AAG CCC GGC CCG GCT CTC AGC TAC CCG CAG GAG GAG        204
Ser Ala Pro Val Lys Pro Gly Pro Ala Leu Ser Tyr Pro Gln Glu Glu
          10                  15                  20

GCC ACC CTC AAT GAG ATG TTC CGC GAG GTT GAG GAA CTG ATG GAG GAC        252
Ala Thr Leu Asn Glu Met Phe Arg Glu Val Glu Glu Leu Met Glu Asp
      25                  30                  35

ACG CAG CAC AAA TTG CGC AGC GCG GTG GAA GAG ATG GAG GCA GAA GAA        300
Thr Gln His Lys Leu Arg Ser Ala Val Glu Glu Met Glu Ala Glu Glu
  40                  45                  50                  55

GCT GCT GCT AAA GCA TCA TCA GAA GTG AAC CTG GCA AAC TTA CCT CCC        348
Ala Ala Ala Lys Ala Ser Ser Glu Val Asn Leu Ala Asn Leu Pro Pro
                  60                  65                  70

AGC TAT CAC AAT GAG ACC AAC ACA GAC ACG AAG GTT GGA AAT AAT ACC        396
Ser Tyr His Asn Glu Thr Asn Thr Asp Thr Lys Val Gly Asn Asn Thr
              75                  80                  85

ATC CAT GTG CAC CGA GAA ATT CAC AAG ATA ACC AAC AAC CAG ACT GGA        444
Ile His Val His Arg Glu Ile His Lys Ile Thr Asn Asn Gln Thr Gly
          90                  95                  100

CAA ATG GTC TTT TCA GAG ACA GTT ATC ACA TCT GTG GGA GAC GAA GAA        492
Gln Met Val Phe Ser Glu Thr Val Ile Thr Ser Val Gly Asp Glu Glu
      105                 110                 115

GGC AGA AGG AGC CAC GAG TGC ATC ATC GAC GAG GAC TGT GGG CCC AGC        540
Gly Arg Arg Ser His Glu Cys Ile Ile Asp Glu Asp Cys Gly Pro Ser
  120                 125                 130                 135

ATG TAC TGC CAG TTT GCC AGC TTC CAG TAC ACC TGC CAG CCA TGC CGG        588
Met Tyr Cys Gln Phe Ala Ser Phe Gln Tyr Thr Cys Gln Pro Cys Arg
                  140                 145                 150

GGC CAG AGG ATG CTC TGC ACC CGG GAC AGT GAG TGC TGT GGA GAC CAG        636
Gly Gln Arg Met Leu Cys Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln
              155                 160                 165
```

```
CTG TGT GTC TGG GGT CAC TGC ACC AAA ATG GCC ACC AGG GGC AGC AAT        684
Leu Cys Val Trp Gly His Cys Thr Lys Met Ala Thr Arg Gly Ser Asn
        170                 175                 180

GGG ACC ATC TGT GAC AAC CAG AGG GAC TGC CAG CCG GGG CTG TGC TGT        732
Gly Thr Ile Cys Asp Asn Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys
    185                 190                 195

GCC TTC CAG AGA GGC CTG CTG TTC CCT GTG TGC ACA CCC CTG CCC GTG        780
Ala Phe Gln Arg Gly Leu Leu Phe Pro Val Cys Thr Pro Leu Pro Val
200                 205                 210                 215

GAG GGC GAG CTT TGC CAT GAC CCC GCC AGC CGG CTT CTG GAC CTC ATC        828
Glu Gly Glu Leu Cys His Asp Pro Ala Ser Arg Leu Leu Asp Leu Ile
                220                 225                 230

ACC TGG GAG CTA GAG CCT GAT GGA GCC TTG GAC CGA TGC CCT TGT GCC        876
Thr Trp Glu Leu Glu Pro Asp Gly Ala Leu Asp Arg Cys Pro Cys Ala
            235                 240                 245

AGT GGC CTC CTC TGC CAG CCC CAC AGC CAC AGC CTG GTG TAT GTG TGC        924
Ser Gly Leu Leu Cys Gln Pro His Ser His Ser Leu Val Tyr Val Cys
        250                 255                 260

AAG CCG ACC TTC GTG GGG AGC CGT GAC CAA GAT GGG GAG ATC CTG CTG        972
Lys Pro Thr Phe Val Gly Ser Arg Asp Gln Asp Gly Glu Ile Leu Leu
    265                 270                 275

CCC AGA GAG GTC CCC GAT GAG TAT GAA GTT GGC AGC TTC ATG GAG GAG       1020
Pro Arg Glu Val Pro Asp Glu Tyr Glu Val Gly Ser Phe Met Glu Glu
280                 285                 290                 295

GTG CGC CAG GAG CTG GAG GAC CTG GAG AGG AGC CTG ACT GAA GAG ATG       1068
Val Arg Gln Glu Leu Glu Asp Leu Glu Arg Ser Leu Thr Glu Glu Met
                300                 305                 310

GCG CTG GGG GAG CCT GCG GCT GCC GCC GCT GCA CTG CTG GGA GGG GAA       1116
Ala Leu Gly Glu Pro Ala Ala Ala Ala Ala Ala Leu Leu Gly Gly Glu
            315                 320                 325

GAG ATT TAGATCTGGA CCAGGCTGTG GGTAGATGTG CAATAGAAAT AGCTAATTTA        1172
Glu Ile

TTTCCCCAGG TGTGTGCTTT AGGCGTGGGC TGACCAGGCT TCTTCCTACA TCTTCTTCCC     1232

AGTAAGTTTC CCCTCTGGCT TGACAGCATG AGGTGTTGTG CATTTGTTCA GCTCCCCCAG     1292

GCTGTTCTCC AGGCTTCACA GTCTGGTGCT TGGGAGAGTC AGGCAGGGTT AAACTGCAGG     1352

AGCAGTTTGC CACCCCTGTC CAGATTATTG GCTGCTTTGC CTCTACCAGT TGGCAGACAG     1412

CCGTTTGTTC TACATGGCTT TGATAATTGT TTGAGGGGAG GAGATGGAAA CAATGTGGAG     1472

TCTCCCTCTG ATTGGTTTTG GGGAAATGTG GAGAAGAGTG CCCTGCTTTG CAAACATCAA     1532

CCTGGCAAAA ATGCAACAAA TGAATTTTCC ACGCAGTTCT TTCCATGGGC ATAGGTAAGC     1592

TGTGCCTTCA GCTGTTGCAG ATGAAATGTT CTGTTCACCC TGCATTACAT GTGTTTATTC     1652

ATCCAGCAGT GTTGCTCAGC TCCTACCTCT GTGCCAGGGC AGCATTTTCA TATCCAAGAT     1712

CAATTCCCTC TCTCAGCACA GCCTGGGGAG GGGGTCATTG TTCTCCTCGT CCATCAGGGA     1772
```

TCTCAGAGGC TCAGAGACTG CAAGCTGCTT GCCCAAGTCA CACAGCTAGT GAAGACCAGA 1832

GCAGTTTCAT CTGGTTGTGA CTCTAAGCTC AGTGCTCTCT CCACTACCCC ACACCAGCCT 1892

TGGTGCCACC AAAAGTGCTC CCCAAAAGGA AGGAGAATGG GATTTTTCTT TTGAGGCATG 1952

CACATCTGGA ATTAAGGTCA AACTAATTCT CACATCCCTC TAAAAGTAAA CTACTGTTAG 2012

GAACAGCAGT GTTCTCACAG TGTGGGGCAG CCGTCCTTCT AATGAAGACA ATGATATTGA 2072

CACTGTCCCT CTTTGGCAGT TGCATTAGTA ACTTTGAAAG GTATATGACT GAGCGTAGCA 2132

TACAGGTTAA CCTGCAGAAA CAGTACTTAG GTAATTGTAG GGCGAGGATT ATAAATGAAA 2192

TTTGCAAAAT CACTTAGCAG CAACTGAAGA CAATTATCAA CCACGTGGAG AAAATCAAAC 2252

CGAGCAGTGC TGTGTGAAAC ATGGTTGTAA TATGCGACTG CGAACACTGA ACTCTACGCC 2312

ACTCCACAAA TGATGTTTTC AGGTGTCATG GACTGTTGCC ACCATGTATT CATCCAGAGT 2372

TCTTAAAGTT TAAAGTTGCA CATGATTGTA TAAGCATGCT TTCTTTGAGT TTTAAATTAT 2432

GTATAAACAT AAGTTGCATT TAGAAATCAA GCATAAATCA CTTCAACTGC TAAAAAAA 2490

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 350 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Gln Arg Leu Gly Ala Thr Leu Leu Cys Leu Leu Leu Ala Ala Ala
-21 -20                 -15                 -10

Val Pro Thr Ala Pro Ala Pro Ala Pro Thr Ala Thr Ser Ala Pro Val
 -5                   1                   5                  10

Lys Pro Gly Pro Ala Leu Ser Tyr Pro Gln Glu Glu Ala Thr Leu Asn
            15                  20                  25

Glu Met Phe Arg Glu Val Glu Glu Leu Met Glu Asp Thr Gln His Lys
        30                  35                  40

Leu Arg Ser Ala Val Glu Glu Met Glu Ala Glu Glu Ala Ala Ala Lys
    45                  50                  55

Ala Ser Ser Glu Val Asn Leu Ala Asn Leu Pro Pro Ser Tyr His Asn
 60                  65                  70                  75

Glu Thr Asn Thr Asp Thr Lys Val Gly Asn Asn Thr Ile His Val His
                80                  85                  90

Arg Glu Ile His Lys Ile Thr Asn Asn Gln Thr Gly Gln Met Val Phe
            95                  100                 105

Ser Glu Thr Val Ile Thr Ser Val Gly Asp Glu Glu Gly Arg Arg Ser
        110                 115                 120
```

```
His Glu Cys Ile Ile Asp Glu Asp Cys Gly Pro Ser Met Tyr Cys Gln
    125                 130                 135

Phe Ala Ser Phe Gln Tyr Thr Cys Gln Pro Cys Arg Gly Gln Arg Met
140                 145                 150                 155

Leu Cys Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln Leu Cys Val Trp
                160                 165                 170

Gly His Cys Thr Lys Met Ala Thr Arg Gly Ser Asn Gly Thr Ile Cys
            175                 180                 185

Asp Asn Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg
        190                 195                 200

Gly Leu Leu Phe Pro Val Cys Thr Pro Leu Pro Val Glu Gly Glu Leu
    205                 210                 215

Cys His Asp Pro Ala Ser Arg Leu Leu Asp Leu Ile Thr Trp Glu Leu
220                 225                 230                 235

Glu Pro Asp Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu
                240                 245                 250

Cys Gln Pro His Ser His Ser Leu Val Tyr Val Cys Lys Pro Thr Phe
            255                 260                 265

Val Gly Ser Arg Asp Gln Asp Gly Glu Ile Leu Leu Pro Arg Glu Val
        270                 275                 280

Pro Asp Glu Tyr Glu Val Gly Ser Phe Met Glu Glu Val Arg Gln Glu
    285                 290                 295

Leu Glu Asp Leu Glu Arg Ser Leu Thr Glu Glu Met Ala Leu Gly Glu
300                 305                 310                 315

Pro Ala Ala Ala Ala Ala Ala Leu Leu Gly Gly Glu Glu Ile
                320                 325
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Pro Ala Pro Arg Arg Arg Trp Leu Leu Leu Leu Ala Val Leu Ala Ala
1               5                   10                  15

Leu Cys Cys Ala Ala Ala Gly Ser Gly Gly Arg Arg Arg Ala Ala Ser
            20                  25                  30

Leu Gly Glu Met Leu Arg Glu Val Glu Ala Leu Met Glu Asp Thr Gln
        35                  40                  45
```

His Lys Leu Arg Asn Ala Val Gln Glu Met Glu Ala Glu Glu Glu Gly
50 55 60

Ala Lys Lys Leu Ser Glu Val Asn Phe Glu Asn Leu Pro Pro Thr Tyr
65 70 75 80

His Asn Glu Ser Asn Thr Glu Thr Arg Ile Gly Asn Lys Thr Val Gln
85 90 95

Thr His Gln Glu Ile Asp Lys Val Thr Asp Asn Arg Thr Gly Ser Thr
100 105 110

Ile Phe Ser Glu Thr Ile Ile Thr Ser Ile Lys Gly Gly Glu Asn Lys
115 120 125

Arg Asn His Glu Cys Ile Ile Asp Glu Asp Cys Glu Thr Gly Lys Tyr
130 135 140

Cys Gln Phe Ser Thr Phe Glu Tyr Lys Cys Gln Pro Cys Lys Thr Gln
145 150 155 160

His Thr His Cys Ser Arg Asp Val Glu Cys Cys Gly Asp Gln Leu Cys
165 170 175

Val Trp Gly Glu Cys Arg Lys Ala Thr Ser Arg Gly Glu Asn Gly Thr
180 185 190

Ile Cys Glu Asn Gln His Asp Cys Asn Pro Gly Thr Cys Cys Ala Phe
195 200 205

Gln Lys Glu Leu Leu Phe Pro Val Cys Thr Pro Leu Pro Glu Glu Gly
210 215 220

Glu Pro Cys His Asp Pro Ser Asn Arg Leu Leu Asn Leu Ile Thr Trp
225 230 235 240

Glu Leu Glu Pro Asp Gly Val Leu Glu Arg Cys Pro Cys Ala Ser Gly
245 250 255

Leu Ile Cys Gln Pro Gln Ser Ser His Ser Thr Thr Ser Val Cys Glu
260 265 270

Leu Ser Ser Asn Glu Thr Arg Lys Asn Glu Lys Glu Asp Pro Leu Asn
275 280 285

Met Asp Glu Met Pro Phe Ile Ser Leu Ile Pro Arg Asp Ile Leu Ser
290 295 300

Asp Tyr Glu Glu Ser Ser Val Ile Gln Glu Val Arg Lys Glu Leu Glu
305 310 315 320

Ser Leu Glu Asp Gln Ala Gly Val Lys Ser Glu His Asp Pro Ala His
325 330 335

Asp Leu Phe Leu Gly Asp Glu Ile
340

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGGAGGATCC GCGCCCGCTC CGACGGCG 28

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCCTCTAGAT TAAATCTCTT CCCCTCCCAG CAGT 34

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TGCCGCGGAT CCGCCATCAT GCAGCGGCTT GGGGCCAC 38

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCACAGGTAC CCACAGCCTG GTCCAGATCT AAATCTCTTC CCCTCCCAG 49

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTCTCTAGAC AGATCTAAAT CTCTTCCCCT CCCAG 35

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTCTCTAGAC AGATCTAAGC GTAGTCTGGG ACGTCGTATG GGTAAATCTC TTCCCCTCCC 60

AGCAG 65

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCTGCCGCGG ATCCGCCACC ATGCAGCGGC TTGGGGCCAC C 41

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CACACGCGGA TCCAGATCTA AATCTCTTCC CCTC 34

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AATTCGGCAC GAGCCTGACT GAAGAGATGG CGCTGAGGGA GCCTNCGGCT GCCGCCGTTG 60

CACTGCTGGG AGGGGAAGAG ATTTAGATCT GGACCAGGCT GTGGGTAGAT GTGCAATAGA 120

AATAGCTAAT TTATTTCCCC AGGTGTGTGC TTTAGGCGTG GGCTGACCAG GCTTCTTCCT 180

ACATCTTCTT CCCAGTAAGT TTCCCCNCTG GCTTGACAGC ATGAGGTGTT GTGCATTTTG 240

TTCAGCTCCC CCAGGCTGTT CTCCAGGNTT CACAGTCTGG TGCTTGGGAG AGTNAAGGCA 300

GGGTTAAACT TCAGGAGCAG TTTGCCACCC NTNGTNCNGA TTATTTGGCT TGCTTTNCCN 360

NTACCATTTG CAAAANAGCC GTTT 384

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AATTCGGCAC GAGGTTCCGC GAGGTTGAGG AACTGATGGA GGACACGCAG CACAAATTGC 60

GCACGCGGTG GAAGAGATGG AGGCAGAAGA AGCTGCTGCT AAAGCATCAT CAGAAGTGAA 120

CCTGGCAAAC TTACCTCCCA GCTATCACAA TGAGACCAAC ACAGACACGA AGGTTGGAAA 180

TAATACCATC CATGTGCACC GAGAAATTCA CAAGATAACC AACAACCAGA CTTGACAAAT 240

GGTCTTTTTC AGAGACAGTT TNACATCTTT GGGAGACGAG AAGCAGAGGN GCNCGNTNCN 300

TATNGCGNGG CTTTTGGCCA GATTNCTNCC ATTTNCAGTT CCNTAACTNC ACCTNCCGGC 360

CANGGTNTTT ACCCGGCATN GTTTTTGGCC ACTTTNTTTG GTATNNCCAA TGCCCNGGAG 420

ATNGCCTTTN NACANGGNTC ACCGGTTTTT TNTTCAAGGG TTTTCTTTTA AATCCTGGGG 480

GGTTCTTCCC ACGTTTGNTT TCT 503

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 490 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGCACGAGGG AGGCTTGAGG TGGAAGTGGG GGTCGGGCAC TCTGACCTGG TCGAGGAGGG 60

GCTAGGGTTT GAACCGGGGA CAGAGTCTAG GTGAGCTGGG GCTTGGGAGC TATTAGCGTA 120

GAGGATCCGG GTTCGGTTGC TCTGGCGAGG GCTCCAGCAT CACAGGCGGC GGCTGCGGGC 180

GCANAGCGTA GATGCAGCGG CTTGGGGCCA CCCTGCTGTG CCTGCTGCTG GCGGCGGCGG 240

TCCCCACGGC CCCCGCGCCC GCTCCGACGG CGACCTCGGC TCCAGTCAAG CCCGGCCCGG 300

CTCTGGACTN ACCCGCAGAG GGANGCCAAC CTNCAATGGA AATGTTTCCG CGNAGTTTGG 360

AGGAATNGAT GGGAAGGACA CGCNANNANA AATTGCGCNA GCGGTTGGGA AGAGATGGAA 420

GGCAAGAAAG AAGTTGCTGG TGAAAGNATC ATCAAGAAAT GGAACTTGGC AAATTGAACT 480

CCCCANNANT 490

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 84 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGCGACCTCG GCTCCAGTCA AGCCCGGCCC GNGNTCTCAG CTACCCATAG GTGGAGGNCA 60

CCCTGNGTGC ANACCCTTGC CCAA 84

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAATGTGGAG TCTCCCTCTG ATTGGTTTTN GGGAAATGTG GAGAAGAGTG CCCTGCTTTG 60

CAAACATCAA CCTGGCAAAA ATGCAACAAA TGNATTTTCC ACGCATTCTT TCCATGGGCA 120

TAGGTAAGCT GTGCCTTCAG CTGTTGCAGA TGAAATGTTC TGTTCACCCT GCATTACATG 180

TGTTTATTCA TCCAGCAGTG TTGCTCAGCT CCTACCTCTG T 221

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 557 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCCACGAGTG CATCATCGAC GAGGACTGTG GGCCCAGCAT GTACTGCCAG TTTGCCAGCT 60

TCCAGTACAC CTGCCAGCCA TGCCGGGGCC AGAGGATGCT CTGCACCCGG GACAGTGAGT 120

GCTGTGGAGA CCAGCTGTGT GTCTGGGGTC ACTGCACCAA AATGGCCACC AGGGGCAGCA 180

ATGGGACCAT CTGTGACAAC CAGAGGGACT GCCAGCCGGG GCTGTGCTGT GCCTTCCAGA 240

GAGGCCTGCT GTTCCCTGTG TGCACACCCC TGCCCGTGGA GGNGAGCTTT GCCATGACCC 300

CGCCAGCCGG CTTCTGGACC TCATCACCTG GGAGCTAGAG CCTGATGGAG CCTTGGACCG 360

ATGCCCTTGT GCCAGTGGCC TCCTCTGCCA GCCCCACAGC CACAGCCTGG TGTATGTGTG 420

CAAGCCGACC TTCGTGGGGA GCCGTGACCA AGATGGGGAG ATCTGCTGCC CAGAGAGGTC 480

CCGATGAGTA TGAAGTTGGA ACTTCATGGA GGAGGTNCGC AAGAACTTGA AGACTTGAGA 540

GGAGCTTACT GAANAAT 557

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 410 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGGACCATCT GTGACAACCA GAGGGACTGC CAGCCGGGGC TGTGCTGTGC CTTCCAGAGA 60

GGCCTGCTGT TCCCTGTGTG CACACCCCTG CCCGTGGAGG GCGAGCTTTG CCATGACCCC 120

GCCAGCCGGC TTCTGGACCT CATCACCTGG GAGCTAGAGC CTGATGGAGC CTTGGACCGA 180

TGCCCTTGTG CCAGTGGCCT CCTCTGCCAG CCCCACAGCC ACAGCCTGGT GTATGTGTGC 240

AAGCCGACCT TCGTGGGGAG CCGTGACCAA GATGGGGAGA TCCTGCTGCC CAGAGAGGTC 300

CCCGATGAGT ATGAAGTTGG CAGCTTCATG GAGGAGGTGC GCCAGGAGCT GGAGGACCTG 360

GGAGAGGAGC CTTGACTTNA AGAGATGGCG CTGAGGGAGC CTTCGGGTTG 410

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 397 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

NAAGTTGGCA GCTTCATGGA GGAGGTGCGC CAGGAGCTGG AGGACCTGGA GAGGAGCCTG 60

ACTGAAGAGA TGGCGCTGGG GGAGCCTGCG CTGCCGCCTT GGCANTGCTG GGAGGGGAAG 120

AGATTTAGAT CTGGACCAGG CTGTGGGTAG ATGTGCAATA GAAATAGCTA ATTTATTTCC 180

CCAGGTNTGT GCTTTAGGCG TGGGCTGACC AGGCTTCTTC CNACATCTTC TTCCCAGTAA 240

GTTTCCCCTC TGGCTTGACA GCATGAGGTG TTNTGCATTT GTTCAGCTCC CCCAGGCTGT 300

TCTCCAGGCT TCACAGTCTT GTGCTTGGGA GAGTCAGGCA GGGTTAAACT GCAGGAGCAG 360

TTTGCCACCC CTGTCCAGAT TATTTGGCTG CTTTGCC 397

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 356 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TGCATCATCG ACGAGGACTG TGGGCCCAGC ATGTACTGCC AGTTTGCCAG CTTCCAGTAC 60

ACCTGCCAGC CATGCCGGGG CCAGAGGATG CTCTGCACCC GGGACAGTGA GTGCTGTGGA 120

GACCAGCTGT GTGTCTGGGG TCACTGCACC AAAATGGCCA CCAGGGGCAG CAATGGGACC 180

ATCTGTGACA ACCAGAGGGA CTNCCAGCCG GGGCTGTGCT GTGCCTTCCA GAGAGGCCTG 240

CTGTTCCCTG TGTGCACANC CCTGCCCGTG GAGGGCGAGC TTTGCCATGA CCCCGNCAGC 300

CGGNTTCTGG ACCTCATCAA CTGGGAGCTA GAGCCTGATG GAGCCTTGGA CCGATG 356

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 319 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CTGGNGAGGA GCCTGACTGA AGAGATGGCG CTGGGGGAGC CTGCGGCTGC CGCCGCTGCA 60

CTGCTGGGAG GGGAAGAGAT TTAGATCTGG ACCAGGCTGT NGGTAGATGT GCAATAGAAA 120

TAGCTAATTT ATTTCCCCAG GTGTGTGCTT TAGGCGTGGG CTGACCAGTC TTCTTCCTAC 180

ATCTTCTTCC CANTAAGTTT CCCCTCTGGC TTGACAGCAT GAGGTGTTGT GCATTTTTTC 240

AGCTCCCCCA GGCTGTTCTC CAGGCTTCAC AGTCTGGTGC TTGGGAGAGT CAGGCAGGGT 300

TAAACTNCAG GAGCAGTTT 319

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 298 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTTCATGGAG GAGGTGCGCC AGGAGCTGGA GGACCTGGAG AGGAGCCTGA CTGAAGAGAT 60

GGCGCTGGGG GAGCCTGCGG CTGCCGCCGT GNCACTGCTG GGAGGGGAAG AGATTTAGAT 120

CTGGACCAGG CTGTGGGTAG ATGTGCAATA GAAATAGCTA ATTTATTTCC CCAGGTGTGT 180

GCTTTAGGCG TGGGCTGACC AGGCTTCTTN CTACATCTTC TTCCCAGTAA GTTTCCCCTC 240

TGGCTTGACA GCATGAGGTG TTGTGCATTT GTTCAGCTCC CCCAGGCTGT TCTCCAGG 298

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 302 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTCAAGCCCG GCCCGGCTCT CAGCTACCCG CAGAGGAGGC CACCCTCAAT GAGATGTTCC 60

GCGAGGTTGA GGAACTGATG GAGGACACGC AGCACAANTT GCGCANGCGG TTGGAAGAGA 120

TGGAGGCAGA AGAAGCTGCT GCTAAAGCAT CATCAGAAGT GAACCTGGCA AACTTACCTC 180

CCAGCTATCA CAATGAGACC AACACAGACA CGAAGGTTGG AAATAATACC ATCCATGTGC 240

ACCGAGAAAT TCACAAGATA ACCAACAACC AGACTGGACA AATGGTCTTT TCAGAGACAG 300

TT 302

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 279 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAGGAGCCTG ACTGAAGAGA TGGCGCTGAG GGAGCCTGCG GCTGCCGCCG TNGCACTGCT 60

GGGAGGGGAA GAGATTTAGA TCTGGACCAG GCTGTGGGTA GATGTGCAAT AGAAATAGCT 120

AATTTATTTC CCCAGGTGTG TGCTTTAGGC GTGGGCTGAN CAGGCTTCTT NCTACATCTT 180

CTTGCCAGTA NGNTTCCCCT CTGGCTTGAC AGCATGAGGT GTTGTGCATT TGTTCAGCTC 240

CCCCAGGCTG TTCTCCAGGC TTCACAGTCT GGTGCTTGG 279

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TACCATCCAT GTGCACCGAG AAATTCACAA GATAACCAAC AACCAGACTG GACAAATGGT 60

CTTTTCAGAG ACAGTTATCA CATCTNTGGG AGACGAAGAA GGCAGAAGGA GCCACGAGTG 120

CATCATCGAC GAGGACTNTG GGCCCAGCAT GTACTGCCAG TTTGCCAGCT TCCAGTACAC 180

CTGCCAGCCA TGCCGGGGCC AGAGGATGCT CTNCACCCGG GACAGTGAGT GCTGTGGAGA 240

CCAGCTGTGT GTCTGGGGTC ACT 263

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGCTGCGGCG CAAGCGANGA TGCAGCGGCT TGGGGCCACC CTGCTGTGCC TGCTGCTGGC 60

GGCGGCGGTC CCCACGGCCC CCGCGCCCGC TCCGACGGCG ACCTCGGCTC CAGTCAAGCC 120

CGGCCCGGCT CTCAGCTACC GCGCAGGAGG AGGCCACCCT CAATGAGATG TTCCGCGAGG 180

TTGAGGAACT GATGGAGGAC ACGCAGCACA AATTGGCACC GGTGGAAGAG ATGGAGGCAG 240

AAGAAGCTGC TGCTAAAGCA TCATCAGAAG TGAACCTGGC AAACTTACCT CCCAGCTATC 300

ACAATGAGAC CAACACAGAC ACGAAGGTTG GAAATAATAC CATCCATGTG CACCGAGAA 359

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ACCAGAGGGA CTGCCAGCCG GGGTGTGCTC GTGCCTTCCA GAGAGGCCTG CTGTTCCCTG 60

TGTGCACACC CCTGCCCGTG GAGCGGACGC TTTGCATGAC CCCGCCAGCC GGCTTCTGGA 120

CCTCATCACC TGGGAGCTAG AGCCTGATGG AGCCTTGGAC CGATGCCCTT GTGCCAGTGG 180

CTCCTCTGCC AGCCCCACAG CCACAGCCTG GTGTATGTGT GCAAGCCGAC CTTCGTGGGG 240

AGCCGTGACC AAGATGGGGA GATCCTGCTG CCCAANAAAG GTCCCCGATT GAGTATGAAG 300

TTGGCAAGCT TCATGGAAGG AANGG 325

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GAGAAATTCA CAAGATAACC AACAACCAGA CTGGACAAAT GGTCTTTTCA GAGACAGTTA 60

TCACATCTGT GGGAGACGAA GAAGGCAGAA GGAGCCACGA GTGCATCATC GACGAGGACT 120

NTGGGCCCAG CATGTACTGC CAGTTTNCCA GCTTCCAGTA CACCTGCCAG CCATGCCGGG 180

GCCAGAGGAT GCTCTGCACC CGGGACAGTG AGTGCTGTGG AGACCAGCTG TGTGTCTG 238

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 236 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TTGGAAATAA TACCATCCAT GTGCACCGAG AAATTCACAA GATAACCAAC AACCAGACTG 60

GACAAATGGT CTTTTCAGAG ACAGTTATCA CATCTGTGGG AGACGAAGAA GGCAGAAGGN 120

GCCACGAGTG CATCATCGAC GAGGACTGTG GGCCCAGCAT GTACTGCCAG TTTGCCAGCT 180

TCCAGTACAC CTGCCAGCCA TGTNGGGGCC AGAGGATGCT CTGCACCCGG GACAGT 236

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CCAGCTGTGT GTCTGGGGTC ACTGCACCAA AATGGCCACC AGGGGCAGCA ATGGGACCAT 60

CTGTGACAAC CAGAGGGACT GCCAGCCGGG GCTGTGCTGT GCCTTCCAGA GAGGCCTGCT 120

GTTCCCTGTG TGCACACCCC TGCCCGTGGA GGGANGCTTT GCCATGACCC CGCCAGCCGG 180

CTTCTGGACC TCATCACCTG GGGAGCTAGA GCCTGATGGA GCCTTGGGAC CGATGCCCTT 240

GTGCCAGTGG CCTCCTCTTG CCAGCCCCAC AGCCACAGCC TGGGTGTATG TTGTGCAAAG 300

CCGACCTTCG TNGGGGAACC GTGACCAAGA TGGGGGAGAT TCTT 344

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTTTTTGGGG AAATAAATTA GCTATTTCTA TTGCACATCT ACCCACAGCC TGGTCCAGAT 60

CTAAATCTCT TCCCCTCCCA GCAGTGCAGC GGCGGCANAG GNCTCCCCCA GCGCCATCTC 120

TTCAGTCAGG CTCCTCTCCA GGTCCTCCAG CTCCTGGCGC ACCTCCTCCA TGAAGCTGCC 180

AACTTCATAC TCATCGGGGA CCTCTCTGGG CAGCAGGA 218

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 247 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCAGACGGAG ATGCAGCGGC GTTGGGGCCA CCCTGACTGT GCCTGCTGCT GGCGGCGGCG 60

GTCCCCACGG CCCCCGCGCC CGCTCCGACG GCGACCTCGG CTCCAGTCAA GCCCGGCCCG 120

GCTCTCAGCT ACCCGCAGGA GCGAGGCCAC CCTCAATGAG ATGTTCCGCG AGGTTGAGGA 180

ACTGATGGAG GACACGCAGC ACAAATTGCG CAGCGGTGGG AAGAGATGGA GGCAGAAGAA 240

GCTGCTG 247

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 210 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTGGNAGAAG GAGCCTGACT GAAAGAGATG GCGCTGGGGG AGCCTGCGGC TGCCGCCGTG 60

NCACTGCTGG GAGGGGAAGA GATTTAGATC TGGACCAGGC TGTGGGTAGA TGTGCAATAG 120

AAATAGCTAA TTTATTTCCC CAGGTGTGTG CTTTAGGCGT GGGCTGACCA GGNTTCTTCC 180

TACATCTTCT TCCCAGTAAG TTTCCCCTCT 210

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CCAGTACTGG TACAATATGG ATCTTTTCAG AGACAGGTTA TCACATCTGT NGGAGACGAA 60

GAAGGCAGAA GGAGCCACGA GTGCATCATC GACGAGGACT GTGGGCCCGG CTCTCAGCTA 120

CCCGCAGAGG AGGCCACCCT CCTHTAGATG TTCCGCGAGT TGAGGACTGA TGGAGGACAC 180

GCTGCACTGC TGGGAGGGGA AGAGATTTAG ATCTGGACCA GGCTGTGGGT AGATGTGCAA 240

TAGAAATAGC TAATTTATTT CCCAGGTGTG TGCTTTAGGC GTGGCTGACC AGGTTCTTCT 300

ACA 303

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CGCTGCACTG CTGGGAGGGG AAGAGATTTA GATCTGGACC AGGCTGTGGG TAGATGTGCA 60

ATAGAAATAG CTAATTTATT TCCCCAGGTG TGTGCTTTAG GCGTGGGCTG CCCAGGCTTC 120

TTCCTACATN TCCGTCCCNG TAAGTTTCCC CTCTAGCGAA AACAGAATAA GGTG 174

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 151 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GACGAAGAAG GCAGAAGGAG CCACGAGTGC ATCATCGACG AGGACTGTGG GCCAAGCATG 60

TACTGCCAGT TTAACAGCTA ACAGTACCAC CTGCCAGCCA TGCCGGAAAA AGAGGATGAC 120

TCTGCACCCG GGACAGTGAG TGACTGTAGG A 151

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 2, line 14

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

AMERICAN TYPE CULTURE COLLECTION

Address of depositary institution *(including postal code and country)*

12301 Parklawn Drive
Rockville, Maryland 20852
United States of America

| Date of deposit | Accession Number |
|---|---|
| September 23, 1996 | ATCC 97728 |

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

DNA plasmid HHFH678
In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

| For receiving Office use only | For International Bureau use only |
|---|---|
| ☒ This sheet was received with the international application | ☐ This sheet was received by the International Bureau on: |
| Authorized officer: Elnora Rivers, Paralegal Specialist, IAPD - PCT Operations, (703) 305-3070 | Authorized officer |

Form PCT/RO/134 (July 1992)

## *What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the amino acid sequence at position about -21 to about 329 in SEQ ID NO:2;

(b) a nucleotide sequence encoding the amino acid sequence at position about -20 to about 329 in SEQ ID NO:2;

(c) a nucleotide sequence encoding the amino acid sequence at position about 1 to about 329 in SEQ ID NO:2;

(d) a nucleotide sequence encoding the CESP polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728;

(e) a nucleotide sequence encoding the mature CESP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; and

(f) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), or (e).

2. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence in SEQ ID NO:1.

3. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in SEQ ID NO:1 encoding the CESP polypeptide having the complete amino acid sequence in SEQ ID NO:2.

4. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in SEQ ID NO:1 encoding the mature CESP polypeptide having the amino acid sequence in SEQ ID NO:2.

5. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in ATCC Deposit No. 97728.

6. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence encoding the CESP polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728.

7. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence encoding the mature CESP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728.

8. An isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide having a nucleotide sequence identical to a nucleotide sequence in (a), (b), (c), (d), or (e) of claim 1 wherein said polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

9. An isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a CESP polypeptide having an amino acid sequence in (a), (b), (c), (d), or (e) of claim 1.

10. The isolated nucleic acid molecule of claim 9, which encodes an epitope-bearing portion of a CESP polypeptide selected from the group consisting of: a polypeptide comprising amino acid residues from about amino acid -1 to about 65 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 71 to about 105 in SEQ ID NO:2; a polypeptide comprising amino acid

residues from about 114 to about 136 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 148 to about 169 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 174 to about 198 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 213 to about 229 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 234 to about 253 in SEQ ID NO:2; and a polypeptide comprising amino acid residues from about 267 to about 315 in SEQ ID NO:2.

11. An isolated nucleic acid molecule, comprising a polynucleotide having a sequence selected from the group consisting of:

(a) the nucleotide sequence of a fragment, wherein said fragment comprises at least 50 contiguous nucleotides of the coding region of the sequence shown in SEQ ID NO:1, provided that said isolated nucleic acid molecule is not SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, or SEQ ID NO:36, or any subfragment thereof; and

(b) a nucleotide sequence complementary to a nucleotide sequence in (a).

12. A method for making a recombinant vector comprising inserting an isolated nucleic acid molecule of claim 1 into a vector.

13. A recombinant vector produced by the method of claim 12.

14. A method of making a recombinant host cell comprising introducing the recombinant vector of claim 13 into a host cell.

15. A recombinant host cell produced by the method of claim 14.

16. A recombinant method for producing a CESP polypeptide, comprising culturing the recombinant host cell of claim 15 under conditions such that said polypeptide is expressed and recovering said polypeptide.

17. An isolated CESP polypeptide having an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) amino acids about -21 to about 329 in SEQ ID NO:2;

(b) amino acids about -20 to about 329 in SEQ ID NO:2;

(c) amino acids about 1 to about 329 in SEQ ID NO:2;

(d) the amino acid sequence of the CESP polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; and

(e) the amino acid sequence of the mature CESP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; and

(f) the amino acid sequence of an epitope-bearing portion of any one of the polypeptides of (a), (b), (c), (d), or (e).

18. An isolated polypeptide comprising an epitope-bearing portion of the CESP protein, wherein said portion is selected from the group consisting of: a polypeptide comprising amino acid residues from about amino acid -1 to about 65 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 71 to about 105 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 114 to about 136 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 148 to about 169 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 174 to about 198 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 213 to about 229 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 234 to about 253 in

SEQ ID NO:2; and a polypeptide comprising amino acid residues from about 267 to about 315 in SEQ ID NO:2.

19. An isolated nucleic acid molecule comprising a polynucleotide encoding a CESP polypeptide wherein, except for one to fifty conservative amino acid substitutions, said polypeptide has a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the amino acid sequence at position about -21 to about 329 in SEQ ID NO:2;

(b) a nucleotide sequence encoding the amino acid sequence at position about -20 to about 329 in SEQ ID NO:2;

(c) a nucleotide sequence encoding the amino acid sequence at position about 1 to about 329 in SEQ ID NO:2;

(d) a nucleotide sequence encoding the CESP polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728;

(e) a nucleotide sequence encoding the mature CESP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; and

(f) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), or (e).

20. An isolated CESP polypeptide wherein, except for one to fifty conservative amino acid substitutions, said polypeptide has a sequence selected from the group consisting of:

(a) amino acids about -21 to about 329 in SEQ ID NO:2;

(b) amino acids about -20 to about 329 in SEQ ID NO:2;

(c) amino acids about 1 to about 329 in SEQ ID NO:2;

(d) the amino acid sequence of the CESP polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; and

(e) the amino acid sequence of the mature CESP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97728; and

(f) the amino acid sequence of an epitope-bearing portion of any one of the polypeptides of (a), (b), (c), (d), or (e).

```
          10                  30                  50
-72  AGAGGATCCGGGTTCGGTTGCTCTGGCGAGGGCTCCAGCATCACAGGCGGCGGCTGCGGG  -13

          70                  90                  110
-12  CGCAGAGCGGAGATGCAGCGGCTTGGGGCCACCCTGCTGTGCCTGCTGCTGGCGGCGGCG   47
 -3              M  Q  R  L  G  A  T  L  L  C  L  L  L  A  A  A     16

         130                 150                 170
 48  GTCCCCACGGCCCCCGCGCCCGCTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCG  107
 17  V  P  T  A  P  A  P  A  P  T  A  T  S  A  P  V  K  P  G  P    36

     190                 210                 230
108  GCTCTCAGCTACCCGCAGGAGGAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAA  167
 37  A  L  S  Y  P  Q  E  E  A  T  L  N  E  M  F  R  E  V  E  E    56

         250                 270                 290
168  CTGATGGAGGACACGCAGCACAAATTGCGCAGCGCGGTGGAAGAGATGGAGGCAGAAGAA  227
 57  L  M  E  D  T  Q  H  K  L  R  S  A  V  E  E  M  E  A  E  E    76

         310                 330                 350
228  GCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGCAAACTTACCTCCCAGCTATCACAAT  287
 77  A  A  A  K  A  S  S  E  V  N  L  A  N  L  P  P  S  Y  H  N    96

         370                 390                 410
288  GAGACCAACACAGACACGAAGGTTGGAAATAATACCATCCATGTGCACCGAGAAATTCAC  347
 97  E  T  N  T  D  T  K  V  G  N  N  T  I  H  V  H  R  E  I  H   116

         430                 450                 470
348  AAGATAACCAACAACCAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTG  407
117  K  I  T  N  N  Q  T  G  Q  M  V  F  S  E  T  V  I  T  S  V   136

         490                 510                 530
408  GGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGTGGGCCCAGC  467
137  G  D  E  E  G  R  R  S  H  E  C  I  I  D  E  D  C  G  P  S   156

         550                 570                 590
468  ATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATGCCGGGGCCAGAGGATG  527
157  M  Y  C  Q  F  A  S  F  Q  Y  T  C  Q  P  C  R  G  Q  R  M   176

         610                 630                 650
528  CTCTGCACCCGGGACAGTGAGTGCTGTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACC  587
177  L  C  T  R  D  S  E  C  C  G  D  Q  L  C  V  W  G  H  C  T   196

         670                 690                 710
588  AAAATGGCCACCAGGGGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCG  647
197  K  M  A  T  R  G  S  N  G  T  I  C  D  N  Q  R  D  C  Q  P   216

         730                 750                 770
648  GGGCTGTGCTGTGCCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTG  707
217  G  L  C  C  A  F  Q  R  G  L  L  F  P  V  C  T  P  L  P  V   236

         790                 810                 830
708  GAGGGCGAGCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCTGGGAGCTA  767
237  E  G  E  L  C  H  D  P  A  S  R  L  L  D  L  I  T  W  E  L   256

         850                 870                 890
768  GAGCCTGATGGAGCCTTGGACCGATGCCCTTGTGCCAGTGGCCTCCTCTGCCAGCCCCAC  827
257  E  P  D  G  A  L  D  R  C  P  C  A  S  G  L  L  C  Q  P  H   276
```

FIG.1A

```
            910                 930                 950
828  AGCCACAGCCTGGTGTATGTGTGCAAGCCGACCTTCGTGGGGAGCCGTGACCAAGATGGG   887
277  S  H  S  L  V  Y  V  C  K  P  T  F  V  G  S  R  D  Q  D  G    296

            970                 990                1010
888  GAGATCCTGCTGCCCAGAGAGGTCCCCGATGAGTATGAAGTTGGCAGCTTCATGGAGGAG   947
297  E  I  L  L  P  R  E  V  P  D  E  Y  E  V  G  S  F  M  E  E    316

           1030                1050                1070
948  GTGCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAAGAGATGGCGCTGGGGGAG  1007
317  V  R  Q  E  L  E  D  L  E  R  S  L  T  E  E  M  A  L  G  E    336

           1090                1110                1130
1008 CCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGATTTAGATCTGGACCAGGCTG  1067
337   P  A  A  A  A  A  A  L  L  G  G  E  E  I  *                  350

           1150                1170                1190
1068 TGGGTAGATGTGCAATAGAAATAGCTAATTTATTTCCCCAGGTGTGTGCTTTAGGCGTGG  1127

                1210                1230                1250
1128 GCTGACCAGGCTTCTTCCTACATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTTGACAGCA  1187

                1270                1290                1310
1188 TGAGGTGTTGTGCATTTGTTCAGCTCCCCCAGGCTGTTCTCCAGGCTTCACAGTCTGGTG  1247

                1330                1350                1370
1248 CTTGGGAGAGTCAGGCAGGGTTAAACTGCAGGAGCAGTTTGCCACCCCTGTCCAGATTAT  1307

                1390                1410                1430
1308 TGGCTGCTTTGCCTCTACCAGTTGGCAGACAGCCGTTTGTTCTACATGGCTTTGATAATT  1367

                1450                1470                1490
1368 GTTTGAGGGGAGGAGATGGAAACAATGTGGAGTCTCCCTCTGATTGGTTTTGGGGAAATG  1427

                1510                1530                1550
1428 TGGAGAAGAGTGCCCTGCTTTGCAAACATCAACCTGGCAAAAATGCAACAAATGAATTTT  1487

                1570                1590                1610
1488 CCACGCAGTTCTTTCCATGGGCATAGGTAAGCTGTGCCTTCAGCTGTTGCAGATGAAATG  1547

                1630                1650                1670
1548 TTCTGTTCACCCTGCATTACATGTGTTTATTCATCCAGCAGTGTTGCTCAGCTCCTACCT  1607

                1690                1710                1730
1608 CTGTGCCAGGGCAGCATTTTCATATCCAAGATCAATTCCCTCTCTCAGCACAGCCTGGGG  1667

                1750                1770                1790
1668 AGGGGGTCATTGTTCTCCTCGTCCATCAGGGATCTCAGAGGCTCAGAGACTGCAAGCTGC  1727

                1810                1830                1850
1728 TTGCCCAAGTCACACAGCTAGTGAAGACCAGAGCAGTTTCATCTGGTTGTGACTCTAAGC  1787

                1870                1890                1910
```

FIG.1B

```
1788  TCAGTGCTCTCTCCACTACCCCACACCAGCCTTGGTGCCACCAAAAGTGCTCCCCAAAAG 1847

                1930                1950                1970
1848  GAAGGAGAATGGGATTTTTCTTTTGAGGCATGCACATCTGGAATTAAGGTCAAACTAATT 1907

                1990                2010                2030
1908  CTCACATCCCTCTAAAAGTAAACTACTGTTAGGAACAGCAGTGTTCTCACAGTGTGGGGC 1967

                2050                2070                2090
1968  AGCCGTCCTTCTAATGAAGACAATGATATTGACACTGTCCCTCTTTGGCAGTTGCATTAG 2027

                2110                2130                2150
2028  TAACTTTGAAAGGTATATGACTGAGCGTAGCATACAGGTTAACCTGCAGAAACAGTACTT 2087

                2170                2190                2210
2088  AGGTAATTGTAGGGCGAGGATTATAAATGAAATTTGCAAAATCACTTAGCAGCAACTGAA 2147

                2230                2250                2270
2148  GACAATTATCAACCACGTGGAGAAAATCAAACCGAGCAGTGCTGTGTGAAACATGGTTGT 2207

                2290                2310                2330
2208  AATATGCGACTGCGAACACTGAACTCTACGCCACTCCACAAATGATGTTTTCAGGTGTCA 2267

                2350                2370                2390
2268  TGGACTGTTGCCACCATGTATTCATCCAGAGTTCTTAAAGTTTAAAGTTGCACATGATTG 2327

                2410                2430                2450
2328  TATAAGCATGCTTTCTTTGAGTTTTAAATTATGTATAAACATAAGTTGCATTTAGAAATC 2387

                2470                2490
2388  AAGCATAAATCACTTCAACTGCTAAAAAAA                              2417
```

FIG.1C

```
 5  GATLLCLLLAAAVPTAPAPAPTATSAPVKPGPALSYPQEEATLNEMFREV  54
    .|.   :|| || .| ..|:.:.::. ::          |.|.||:|||
 7  PAPRRRWLLLLAVLAALCCAAAGSGGRRRA.........ASLGEMLREV  46

55  EELMEDTQHKLRSAVEEMEAEEAAAKASSEVNLANLPPSYHNETNTDTKV  104
    |.||||||||||.||:||||||.:||  ||||:.||||.||||.||:|::
47  EALMEDTQHKLRNAVQEMEAEEEGAKKLSEVNFENLPPTYHNESNTETRI  96

105 GNNTIHVHREIHKITNNQTGQMVFSETVITSVGDEEGRRSHECIIDEDCG  154
    ||.|:::.|.||.|:|:|.::  :||||:|||: ::|.:|.|||||||||:
97  GNKTVQTHQEIDKVTDNRTGSTIFSETIITSIKGGENKRNHECIIDEDCE  146

155 PSMYCQFASFQYTCQPCRGQRMLCTRDSECCGDQLCVWGHCTKMATRGSN  204
    .:.||||..|:|.||||:.|:  |.|| ||||||||||.| | ..||.|
147 TGKYCQFSTFEYKCQPCKTQHTHCSRDVECCGDQLCVWGECRKATSRGEN  196

205 GTICDNQRDCQPGLCCAFQRGLLFPVCTPLPVEGELCHDPASRLLDLITW  254
    ||||:||:||||| |||||::|||||||||| ||| ||||..|||:||||
197 GTICENQHDCNPGTCCAFQKELLFPVCTPLPEEGEPCHDPSNRLLNLITW  246

255 ELEPDGALDRCPCASGLLCQPH.SHSLVYVCKPTFVGSRDQDGE......  297
    ||||||.|:||||||||:|||: ||| . ||. .  :.|..: |
247 ELEPDGVLERCPCASGLICQPQSSHSTTSVCELSSNETRKNEKEDPLNMD  296

298 .....ILLPREVPDEYEVGSFMEEVRQELEDLERSLTEEMALGEPAAAAA  342
        |:||:: .:|| :|.::|||.|||    ||.:: :: .. .:|
297 EMPFISLIPRDILSDYEESSVIQEVRKELE....SLEDQAGVKSEHDPAH  342

343 ALLGGEEI  350
    .|: |:||
343 DLFLGDEI  350
```

FIG.2

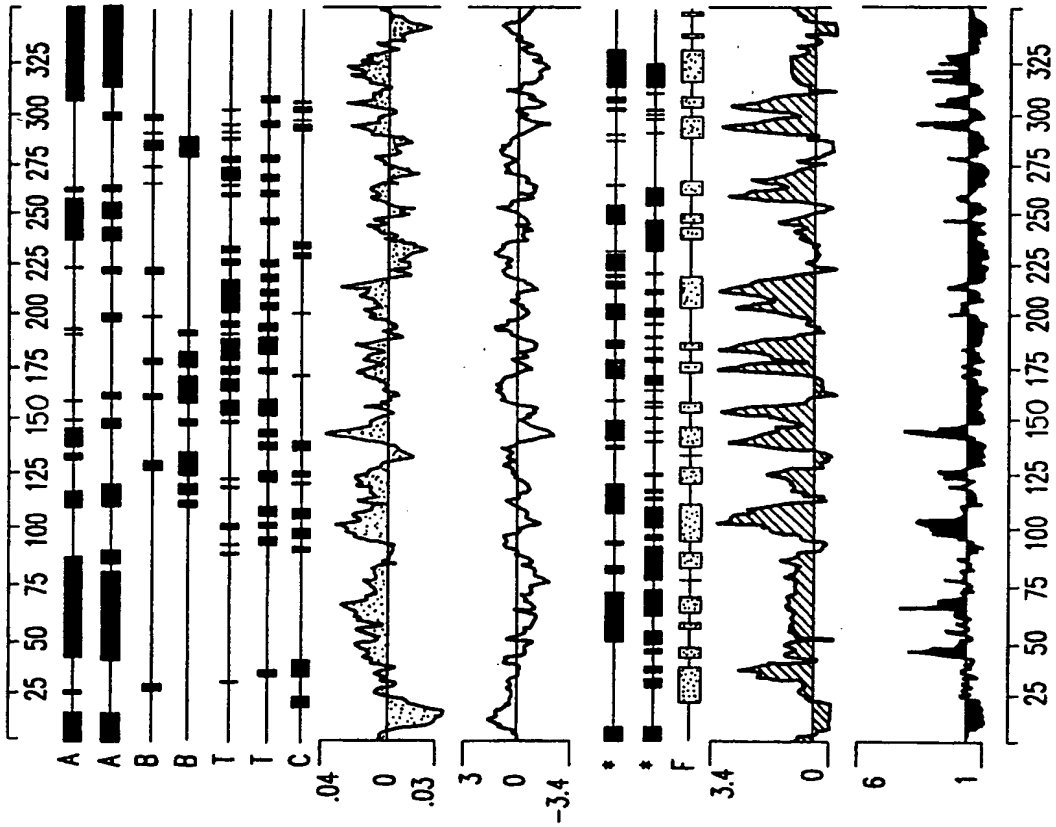
ALPHA, REGIONS-GARNIER-ROBSON
ALPHA, REGIONS-CHOU-FASMAN
BETA, REGIONS-GARNIER-ROBSON
BETA, REGIONS-CHOU-FASMAN
TURN, REGIONS-GARNIER-ROBSON
TURN, REGIONS-CHOU-FASMAN
COIL, REGIONS-GARNIER-ROBSON
HYDROPHILICITY PLOT-KYTE-DOOLITTLE
HYDROPHOBICITY PLOT-HOPP-WOODS
ALPHA, AMPHIPATHIC REGIONS-EISENBERG
BETA, AMPHIPATHIC REGIONS-EISENBERG
FLEXIBLE REGIONS-KARPLUS-SCHULZ
ANTIGENIC INDEX-JAMESON-WOLF
SURFACE PROBABILITY PLOT-EMINI
25 50 75 100 125 150 175 200 225 250 275 300 325

FIG.3

**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/515 — A3

(11) International Publication Number: **WO 98/27932**

(43) International Publication Date: 2 July 1998 (02.07.98)

(21) International Application Number: PCT/US97/23518

(22) International Filing Date: 18 December 1997 (18.12.97)

(30) Priority Data:
60/033,870 20 December 1996 (20.12.96) US

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US).

(74) Agents: STEFFE, Eric, K.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005–3934 (US) et al.

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(88) Date of publication of the international search report:
28 January 1999 (28.01.99)

(54) Title: CEREBELLUM AND EMBRYO SPECIFIC PROTEIN

(57) Abstract

The present invention relates to a novel cerebellum and embryo specific (CESP) protein which is a member of the myocardial factor superfamily. In particular, isolated nucleic acid molecules are provided encoding the human CESP protein. CESP polypeptides are also provided as are vectors, host cells and recombinant methods for producing the same.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# INTERNATIONAL SEARCH REPORT

Interna · l Application No
PCT/US 97/23518

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/515

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | HILLIER ET AL.: "The WashU-Merch EST Project" EMBL DATABASE EMEST9,16 December 1995, XP002084753 EST HS266245;w99c06.s1 Homo sapiens cDNa clone 260362 3' Accession number H99266 compare nucleotides 1-465 from HS266245 with the inverse of nucleotides 2482-2019 in SEQ ID NO:1 --- -/-- | 1,8,11 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 18 November 1998 | 01/12/1998 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016 | Cupido, M |

3

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Interna l Application No
PCT/US 97/23518

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | HILLIER L ET AL.: "The washU-Merck EST Project" EMBL DATABASE EMEST8,27 May 1996, XP002084754 EST HS126340;zc20c09.s1 Soares enescent fibroblasts NbHSF Homo sapiens cDNA clone 322864 3' Accession number W45126 Compare nucleotides 4-433 from HS126340 with the inverse of nucleotides2489-2058 in SEQ ID NO:1 --- | 1,8,11 |
| A | WO 92 06194 A (CONSIGLIO NAZIONALE DELLE RICERCHE) 16 April 1992 see the whole document --- | 1-20 |
| A | MAZUR W ET AL: "Coronary Restenosis and Gene Therapy" TEXAS HEART INSTITUTE JOURNAL, vol. 21, no. 1, 1994, pages 104-111, XP000619173 see the whole document ----- | 1-20 |

3

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Interna Application No
PCT/US 97/23518

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 9206194 A | 16-04-1992 | IT 1242149 B | 16-02-1994 |
| | | AU 8635391 A | 28-04-1992 |
| | | CA 2092533 A | 28-03-1992 |
| | | EP 0550519 A | 14-07-1993 |
| | | JP 6503225 T | 14-04-1994 |